DOCUMENT RESUME

ED 455 263 TM 033 066

AUTHOR	Phelps, Richard P.
TITLE	Test Bashing Series.
PUB DATE	2001-00-00
NOTE	101p.; Some printout pages missing some text at the right margin.
AVAILABLE FROM	For full text: http://www.EducationNews.org.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; *Achievement Tests; Educational Change; Elementary Secondary Education; *Standardized Tests; State Programs; Test Results; *Test Use; Test Wiseness; Testing Problems; *Testing Programs
IDENTIFIERS	*Texas Assessment of Academic Skills

ABSTRACT

This document consists of 14 articles which appeared in the electronic news bulletin, "EducationNews.org," and which were part of a series on "Test Bashing," a discussion of the use of standardized tests that focuses on controversies surrounding the Texas Assessment of Academic Skills and test score improvements in Texas. The articles are: (1) "The 'Washington Post' Advocates State Student Testing Programs, But Not in Texas. Test Bashing Texas, Part 1"; (2) "The Research Sez...Standardized Tests Are Horrible and Terrible. Test Bashing Texas, Part 2"; (3) "The Education Press's Cop-Out on Student Testing. Test Bashing, Part 3"; (4) "That 'Backlash' That Testing Opponents So Desperately Crave. Test Bashing, Part 4"; (5) "More on Texas Testing and School Reform. Test Bashing, Part 5"; (6) "The Education Press's Cop-Out on Testing (continued). Test Bashing, Part 6"; (7) "The Unfortunate Bias of the Education Writers' Association. Test Bashing, Part 7"; (8) "Walt Haney's Texas Obsession, Part 1: SAT Scores. Test Bashing, Part 8"; (9) "Walt Haney's Texas Obsession, Part 2: NAEP Scores. Test Bashing, Part 9"; (10) "Walt Haney's Texas Obsession, Part 3: Texas as a Pariah State. Test Bashing, Part 10"; (11) "Walt Haney's Obsession, Part 4: Dreams of Dropouts. Test Bashing, Part 11"; (12) "Education Week's Anti-Testing Bias: It May Start at the Top. Test Bashing, Part 12"; (13) "The Sad End of Objectivity at the Education Commission of the States. Test Bashing, Part 13"; and (14) "The 'New' Rand Report: Of Course, It's Biased and Contrived, and You're Surprised? Test Bashing, Part 14." (SLD)

ENTIRE DOCUMENT:
POOR PRINT QUALITY

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 455 263

Test Bashing Series

by:

Richard P. Phelps

R. Phelps

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033066

BEST COPY AVAILABLE

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001
Free On-Line Subscription
DailyNews
Commentaries & Reports Bulletin Board Home
Search this site
Over 3,300,000 Visits in 200

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing
Test Bashing Series
Intellectual Property and Digital Information in Higher Education
LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us
Advertising
Archives
DailyNews
Past Five Days
Book Reviews
College Aid
Book Reviews
College Aid

Test Bashing Texas, Part 1:

The *Washington Post* Advocates State Student Testing Programs, But Not in Texas

by Richard P. Phelps

I've lived in Washington for almost a decade and I like it here. As a news information junkie, I could hardly find a better place to live. In addition to the network news most Americans have access to, we have three broadcast public TV stations. Our cable TV lineup has over a dozen primarily news channels in addition to the major network channels. We even have C-Span on radio; can you imagine people in any other town who would want to listen to Congress jabber on all day, day after day? Unlike most U.S. cities anymore, we are also lucky enough to retain 2 major daily newspapers.

The one with the most pretensions, of course, is the *Washington Post*, the paper that broke the Watergate scandal. It envisions itself the country's newspaper of record on political issues, a not unrealistic aspiration given its location.

Until April, I believed what I read in the *Post*. That is, I trusted that its reporters and editors made an effort to check out the reliability and validity of their information, to get all sides to their stories, to achieve some balance in viewpoints. I assumed that mere opinions were confined to the editorial pages. Really, what choice does a reader have but to trust? I don't know much about bauxite mining in New Guinea, so if the Post runs a story on bauxite mining in New Guinea, I'm likely to believe that it's accurate.

I trust no more. In late April, the *Post* published a front-page story on student testing ("An Education 'Miracle' or Mirage?" April 21, pp.A1,4), a topic I do happen to know something about. The *Post* declares itself to be a strong, uncompromising advocate of clear academic standards and the use of high-stakes student tests to enforce those standards. It has supported over the past decade and continues to encourage the efforts of the governors and their states to implement such systems.

This news story, however, was scathing, an unrestrained, wholly critical attack on one of those programs, a particular state testing program that is structured in a way typical of most of the other current state testing programs. The particular state testing program they don't like happens to be in Texas. Oh, by the way, did you know that the governor of Texas is running for president? And, oh, by the way, did you know he belongs to the political party the *Post* tends not to endorse?

The article was certainly one-sided. It included 26 paragraphs attacking the program and allowed 7 paragraphs of defense from state officials and organization spokespersons, stuck mostly in the middle (giving the critics the first and last words). The reporter strongly

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

implied that all testing "experts," by which I suppose he meant academics and other full-time researchers, stand uniformly in opposition to Texas' tests, which is far from the truth.

There are at least several reports, more studious, thorough, and thoughtful than the casual sources the *Post* reporter cited, that have painted the Texas testing program in a positive light, along with the testing program in North Carolina (managed in its entirety by Democratic administrations) that is very similar in its structure and its success. For starters, one report was written by David Grissmer of Rand for the National Goals Panel, another by me for the Fordham Foundation, and another by the Southern Regional Education Board. No mention of any of them in the *Post* article.

But, one doesn't need to read the research on the other side to detect the inconsistencies in the *Post*'s arguments. For example, the first half of the article discussed how difficult the Texas Assessment of Academic Skills (TAAS) is - "drab factories for test preparation," entire instructional budgets spent on "commercial test preparation materials," schools "handed over to 'test prep' from New Year's through April," "TAAS [test] camps," Friday night "lock-ins...where students do TAAS 'drills' until sunup," prizes given to students who do well, and "students cannot graduate if they fail the exams."

The second half of the article then tells us that the test is too easy - "could be passed by many fifth-graders," "'low expectations' are 'cause for concern,'" Which is it? Is the test too difficult or too easy?

Texas' scores on the National Assessment of Educational Progress (NAEP) are used by the *Post*'s reporter (on p.1, para.4) as a benchmark to prove that the "achievement gap" between minority and white students has not, in fact, been narrowing, as scores on the Texas test alone would indicate. Much later, on p.4, para.19, it is mentioned that average NAEP scores for all Texas students have been rising since the Texas testing program was introduced, corroborating that, on average, Texas students are learning more. (Indeed, Texas NAEP scores have risen the most of all the states in the 1990s, a period of high-stakes testing in Texas.) But, the reporter quickly dismissed the NAEP scores in *this* case as insignificant, and again brings up the problem that the "achievement gap" really isn't narrowing as Texas officials say. Which is it? Is the NAEP a valid benchmark or not?

The *Post* reporter also neglects to mention that, regardless of whatever may be happening with his precious "achievement gap," Texas' minority students' scores on the NAEP also have been increasing since the Texas testing program was introduced, along with the overall Texas student average. This suggests that minority students are learning more as a result of the Texas testing program, a concrete accomplishment that will improve their lives.

Six paragraphs in the front of the article were devoted to criticisms from three of the reporter's "experts" before it is mentioned, in a parenthesis in paragraph 28, that all three were paid witnesses in a failed lawsuit to stop the testing program by alleging ethnic bias. The judge didn't find these people's arguments credible, but the Post reporter accepts them as truth. Indeed, all the critics in the article are given the label "expert" whereas none of the test's defenders are. I don't have the time or space needed here to give a fair picture of the manner in which the reporter's favorite "experts" conduct their "research." But, I have written about their objectivity-challenged research methods elsewhere, and I refer the

(1)

Parental Opt-Out Form

Public Policy Org

NEW Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research
The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

reader to those writings.

Perhaps, the reporter was not familiar with the radical egalitarian and radical constructivist philosophies of his preferred "experts." But, he should be, since he parroted much of their terminology. Traditional teaching methods are described variously as: "drudgery," "isolated drills," "going over dreary one-paragraph passages," and "repetitive drills, worksheets, and practice tests." In preparing students for the Texas test, teachers use "aggressive test drilling," and "test coaching," and the test itself is "vulnerable to...teachers' coaching." This latter argument begs the question: does the *Post* want a test so obscure in structure and content that teachers cannot help students prepare for it?

What happened inside the state's schools before the Texas tests were introduced was characterized by the *Post*'s "experts" as wonderful: "creative writing, literature, and science labs," "learning a variety of forms of writing, studying mathematics aimed at problem-solving and conceptual understanding," and "the joy and magic of reading." Radical constructivists label any changes in classroom activity caused by testing as "corruptions" of the natural environment of the classroom and "pollution" of the natural relationship between student and teacher. To their minds, classrooms should operate something like method actors' studios, with absolutely as little structure and organization as is possible.

It was not entertained as possible by the reporter's "experts" that what teachers "naturally" teach in the absence of standards could be less than wonderful. Teachers left to their own devices to create all curriculum from scratch on their own time, because there are no common standards...that's wonderful. The high school coaches who spend as much time talking about their team's progress as the subject matter they are supposed to teach, with the complicity of many of the students...that's wonderful. The English teachers who prefer holding rap sessions on the big ideas in literature or current events instead of teaching writing, because teaching writing is a lot of work...that's wonderful. Teachers teaching subject matter that they happen to like personally because, absent common standards they can teach anything they please whether or not it serves the students' needs...that's wonderful.

Moreover, what happens in the school as a whole, absent common and enforced standards, is presumed to be just as wonderful. Schools that place sports or social achievement higher than academic achievement or that employ social promotion as their chief academic standard, for example, are employing natural, "uncorrupted" structures on their students.

This discussion begs another question: if education in Texas ten years ago was so wonderful, why did the good, level-headed people of Texas collectively decide to put themselves through the enormous hassle and expense of completely overhauling their education system? Why did they put themselves through the laborious, time-consuming process of developing and adopting valid academic standards and fair tests that would enforce them? Were ten million voters and parents duped by some meanies who hate kids and want them to be stupid? Maybe they were duped by ogres from the Texas business community; we all know that business people deep down want to turn all of us into brain dead drones after all, don't we?

Or, could it be that Texas students, particularly the poor and minority students, were not learning as much ten years ago, before the tests were introduced, as the *Post* reporter would have us believe? Perhaps the good citizens of Texas adopted standards and tests

5

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

because "joy and magic" was producing illiterates.

Those standards that the Texas tests measure, by the way, were adopted through a multi-year process that incorporated the input of thousands of citizens. They were adopted through a democratic process, as they are supposed to be. In case the *Washington Post* has not heard, citizens and parents have as much right to set a state's curriculum as do the radical constructivists their reporter likes. Their "experts" had their chance to influence the state curriculum several years ago when Texas was adopting its academic standards. Complaining about them now (the test "does not measure what their students need to learn") is just sour grapes. According to the vast majority of the citizens of Texas, it does.

It's the citizens of Texas, after all, and not a small group of ideological zealots on some of our education school faculty, who get to decide what gets taught in Texas schools. That irritates the zealots to no end, as they think they should be able to decide what other peoples' children are taught. Too bad for them. We do not live in a one-party state where ideological purity is valued over practical, measurable outcomes. The zealots' ideology and irritation are not reasonable bases for setting policy in our system of government.

But, are the majority in Texas leaving poor and minority students behind, as the critics claim? A major argument of the *Post* reporter was that students in lower track classrooms get easier work to do. Most of us would think that this makes perfect sense - students who are not mastering more difficult material are given easier material or taught at a slower pace, with the hope that at least they can master the basic skills. Better for them if they learn how to read and write, even if it means they miss out on dissecting tadpoles.

There is nothing preventing these students from moving on to the allegedly more interesting material, however, once they've mastered the basics. We all know that slower students can catch up if they get extra help (and if they want to do the extra work). One could well argue that social justice demands they be given the opportunity. So, should we not all celebrate a system that guarantees them that extra help so they can catch up? Where do we find such school systems?

In states with high-stakes testing, that's where. In states *without* high-stakes testing, standards don't matter, meeting standards doesn't matter, and students are promoted and graduated no matter how little they learn. There is no need for extra help, after-hours tutoring, summer school, or Saturday classes. In high-stakes testing states, students in academic trouble get the extra help they need. In states without testing, they get forgotten.

Given the rabid and usually dishonest opposition they face from the vested interests in the education research establishment, and the harassment from naive or biased journalists like the *Post*'s, the officials who muster the political courage and considerable effort to develop and put standards in place and then develop and administer tests matched to them, should be greatly admired. There is no more difficult job in our country. Those who take it on should be considered heroes.

In sum, the Post article on testing in Texas was the single worst piece of journalism I've ever read from a media outlet with pretensions for objectivity. But, hey, we all make mistakes. A newspaper can rectify a mistake-prone or one-sided story by printing letters and op-ed pieces that correct the mistakes or present the other side of the story. I sent the *Post* a letter, with similar content to this article's. I also know of others who sent letters.

None were printed. A few days after the story, the Post published an editorial entitled "Don't Flinch on Standards," reiterating their steadfast and uncompromising support of state high-stakes testing programs.

One week after the Texas testing story, the same reporter wrote another *Post* front-page story on state government housing policy in Texas. Did you know that Texas's housing policy is awful and scandalous?

Richard P. Phelps is the author of "Why Testing Experts Hate Testing" (www.edexcellence.net/library/phelps.htm) and other articles on testing one can find on the Web or in journals.

1. "Test Basher Arithmetic," ***Education Week***, March 11, 1998; "The Fractured Marketplace for Standardized Testing," book review, ***Economics of Education Review***, (v.13 n.4) December, 1994; "Estimating the Cost of Systemwide Student Testing in the United States," ***Journal of Education Finance,*** Winter, 2000; ***Why Testing Experts Hate Testing***, Thomas B. Fordham Foundation, Vol.3, No.1, Jan. 1999; "Extent and Character of Systemwide Testing in the U.S." ***Educational Assessment***, V.4, N.2, 1997; "Are U.S. Students the Most Heavily Tested on Earth?" ***Educational Measurement***, V.15, N.3, Fall 1996; "Test Basher Benefit-Cost Analysis," ***Network News & Views***, The Educational Excellence Network, March 1996.
(www.edexcellence.net/issuespl/subject/standar/testbash.html)

Post your comments on this article

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing Texas, Part 2:

The Research Sez...Standardized Tests are Horrible and Terrible!!!

by Richard P. Phelps

The first time, I was surprised. I was reading a research report written by an education prof attacking the use of standardized tests. It was a prototypical example of lying with statistics numbers were left out if they didn't serve the conclusion, data were altered, costs were mad benefits were ignored, numbers were labeled with misleading descriptions, definitions were changed surreptitiously, supportive research was cited that did not contain the evidence clai and solid research that did not support the preferred conclusion was ignored.

I thought it was an exception. Then I read more research reports with doctored data or emp conclusions derived from no data. I realized that I wasn't reading aberrant research; I had encountered the standard research methodology among education professors who hate standardized tests. Nor, I am convinced, was most of it research done incompetently. This research conducted by intelligent people, some considered tops in their field. It was simply

More common than the lying-with-statistics studies, however, were those that use semantic distortions, with only illusions of data.. Yes, I, too, would like to think of a phrase that soun better than "semantic distortions." But, this is what I mean by it:

lots of scientific-sounding language used without any real science behind it;

lots of research-like language used without any valid research behind it;

educational practices that testing opponents like are given euphemistic labels and referred to attractive language, without descriptions of what the practices entail in practical terms; and

educational practices that testing opponents do not like are referred to with unappealing lan without descriptions of what the practices entail in practical terms.

I'm convinced that the reason testing opponents employ euphemisms to describe instruction practices they favor is because they know that if they described in practical terms what they actually mean the public would be appalled. Take for instance their opposition to facts, cont and substance, in favor of process. Most humans believe that intelligence involves both cont process. You can't do data processing without both data and process. You can't speak a lan without both vocabulary and grammar.

Most testing opponents, however, side with the radical constructivists on our education sch faculty and deride content as unimportant to teach, saying you can always look it up. But, h

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

you even begin to know where to look up a word if you know no words? How can you kno where to look up a fact on a computer if you know no facts to begin with?

The testing opponents don't get into this detail. They say we should get rid of standardized t because they promote "lower-order skills," "rote recall," "memorization," "drill and kill," "lo content" and other bad-sounding things. Get rid of tests and teachers can concentrate on "higher-order skills" and "real learning." They hope people won't ask them to explain what " learning" actually means. (By the way, I think they're wrong about standardized tests; I thin test both "lower" and "higher" skills, even in they way testing opponents define the terms.)

E.D. Hirsch has taken on this process-to-the-exclusion-of-content issue in a big way, of cou an appendix to his best-seller *The Schools We Need and Why We Don't Have Them*, he pro enormously useful guide to the jargon and euphemisms used by the radical constructivists. I borrow from his 33-page appendix, "Critical Guide to Educational Terms and Phrases," in t remainder of this article. As a useful public service, The Texas Education Consumers Assoc has abridged Hirsch's appendix and placed it on the World Wide Web, under the title "Educ Terminology Every Parent Must Understand." (http://www.fastlane.net/~eca/Terminology.

Our country owes Professor Hirsch an enormous debt of gratitude for his effort and eruditi raising the linguistic facade that disguises the radical constructivist faith.

Taking up the mantle for the radical constructivists against the Texas Assessment of Acade Skills (TAAS) is a Rice University professor who has been quoted widely in the press natio in recent months. Articles of hers can be found in EducationNews.org.(1) A few years ago t of the country focused on the Texas test because it was so successful. Now, the rest of the focuses on the Texas test because the governor is running for president.

Here's a sampling of the phraseology she uses in criticizing the TAAS, arranged by category some reaction in each.

scientific-sounding language

"We present here our strong assessment..." "Our analysis draws on emerging research..." "O investigations..." "Our investigations..." "Our research required fieldwork in schools and classrooms and frequent interactions with students, teachers, and administrators, whose voi experiences are vital to capture." "...gather and triangulate data from a variety of sources ov multi-year period." "...represent strong, persistent trends emerging from the data." "Our ana reveals..."

I particularly like the cool, research-sounding word "triangulate."

The Rice professor claims that her research is unique to Texas and the TAAS but when first reading her work, I realized that I had read it a dozen times before, written by others. Take word "Texas" out of what she writes and it is a generic radical constructivist's criticism of a standardized test. I doubt that even any of the phrases she uses are original.

Her data, it turns out, consists of nothing more than her talking to some teachers and administrators whose opinions happened to coincide with her own, and the standard researc rhetoric of testing opponents.

Parental Opt-Out Form

Public Policy Org

NEW Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

standardized testing crowds out other topics

"...crowds out other forms of learning..." "This testing system distances the content of curri from the knowledge base of teachers..." "...the TAAS system of testing is reducing the quali quantity of education offered..." "... a regular education has been supplanted by activities w sole purpose is to raise test scores on this particular test." "...fosters an artificial curriculum. "...diverting scarce instructional dollars away from such high quality curricular resources as laboratory supplies and books toward test-prep materials and activities of limited instruction value." "...drilling students on practice exam materials." "...takes time from real teaching." " test scores at the expense of substantive learning."

The crowding-out or narrowing-the-curriculum argument is one of testing opponents' favor and most illogical. The length of the school day was not shortened when the TAAS was introduced; there remains as much time available for instruction as there was before. Grante is taught during the instructional day may be different but, after all, that's the whole point. T citizens of Texas feared that too many fluffy, low-content courses were taking time away fr primary academic subjects.

If testing opponents want to argue that some topics have been "crowded out," in fairness th acknowledge that other topics, considered more important by the citizens of Texas, have be added in. It may be that the selection of courses has been "narrowed" if more time is spent o reading, writing, and arithmetic and less on courses considered peripheral but, at the same ti curricular content of those subject areas considered most important has been broadened, no narrowed.

The Rice professor makes it clear that she doesn't consider any learning that is measured on standardized test to be worthwhile. She also doesn't consider test preparation to be "real lea Workbooks are bad; books are good. That's her opinion. But, there are measurable, demons gains to this type of instruction; that's why it is employed. Responsible teachers and adminis in Texas are using methods they have found to be successful, to insinuate otherwise is sland Conveniently, the Rice professor can argue that her "real learning," the radical constructivis is so individualized that it cannot be measured on a standardized test. We're just supposed t accept her word that it works.

Maybe you were a more conscientious student than I was, but, when I was a kid, I only "rea learned" when I was studying for exams. And, during exams, I often realized concepts for t time or, otherwise, had some knowledge burnt into my brain by the pressure of writing the information that I would have otherwise forgotten.

standardized testing is unfair to minority students

"...this over-reliance on test scores has caused a decline in educational quality for those stud who have the greatest educational need." "The system's popularity is further bolstered by th that it must be improving the education of Latino and African American children, since, in parts of the state, their test scores are also rising." "This testing system distances the conten curriculum from the knowledge base of teachers and from the cultures and intellectual capa the children." "Most damaging are the effects of the TAAS system of testing on poor and m students." "...divorced from children's experience and culture."

Testing opponents don't tell you this, but minorities tend to do relatively better on multiple- standardized tests than on the open-response or performance-based tests that radical

BEST COPY AVAILABLE

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

constructivists favor. The reason probably refers to one of multiple-choice tests' most underappreciated advantages - when a student knows that the correct answer is among the provided, the domain of possible answers is bounded, the type of answer desired is pretty cl With open-ended formats, the domain of possible answers is, well...open ended. The respon doesn't know if the question is looking for a general answer or specific answer, a long answ short answer, a broad answer or a deep, narrow answer. Open response formats may be cult biased.

Just how culturally biased can mathematics be? ...or science?...or geography? Yes, reading comprehension and grammar could be if one doesn't insist that the test be based on Standar English but, if it is based on Standard English, how culturally biased can it be? Even history; there are agreed-upon state standards, and teachers teach to those standards, every student been exposed to the same material. You can't tell me that a test is biased because it asks a q about Hildegard von Bingen and because she was white. It's a rare white kid who knows ab Hildegard without learning it in school.

Got too much on your plate?

As for the "intellectual capacities of the students," the nastiest part of the radical constructiv opposition to standardized tests is the way they use minorities. First, they claim to defend th their culture, declaring that we must teach different subject matter in every school, subject that is tuned to their culture. Conveniently, we could not then use standardized tests to mea school performance because the subject matter wouldn't be standardized. Second, they clai we cannot have high academic standards because minorities won't be able to meet them. An why won't minorities be able to meet them? Are minorities for some reason incapable of me high academic standards? Who's being bigoted?

it's low-quality instruction with standardized tests, not "real learning"

"Much of the drill time is spent learning how to bubble in answers, how to weed out obviou wrong answers, and how to become accustomed to multiple-choice, computer-scored form

"...drilling students on practice exam materials." "...takes time from real teaching."

"...raise test scores at the expense of substantive learning." "...the TAAS system of testing is reducing the quality and quantity of education offered..." "... a regular education has been supplanted by activities whose sole purpose is to raise test scores on this particular test."

"...fosters an artificial curriculum..." "...not a curriculum that will educate these children for productive futures..." "...diverting scarce instructional dollars toward test-prep materials an activities of limited instructional value." "...aimed at the lowest level of skills and informatio

Some criticism of standardized testing is just silly. How long can it take to teach a student h fill in a circle on an answer sheet?

All measures that compare use a structure. There are common scales used to weigh highwa trucks, for example, and all get charged the same rate. There are common rules to any sport contest, and those who know the rules will do better than those who don't know the rules, everything else being equal. So, everyone should be made to know the rules and the rules sh be applied the same way to everyone. It is the standard set of rules, and the uniform applica those rules, that make a measurement system fair. Standardized multiple-choice tests are mu much fairer measures of student academic achievement than the kind of measures the radica

constructivists would like to use.

Yes, it does take some time to learn the rules, but not that much time. Learning the structur standardized test is far simpler than learning successful strategies for some of the board or v games our children play routinely.

Testing opponents would have you believe that one can do well on a multiple-choice test wi even reading the questions, just strategizing over the list of answers. One or two responses "obviously" be wrong. How can one know that some answers are "obviously" wrong witho knowing something about the topic? It is true, that one can get partial credit on standardize A student may know enough to realize that the correct answer has to be b or c, and cannot or e, but doesn't know which and so guesses. A student will get some of these types of ans right at a higher rate than she would just by guessing blindly. That's how one can get partial for knowing part of the answer.

Radical constructivists are big fans of giving partial credit; so why would they dislike this fe of multiple-choice tests? Or, using process-of-elimination to choose among possible answer use process-of-elimination reasoning in our lives all the time; it's a useful, valid method for s problems, both practical and otherwise. You know the butler did it because you eliminated t possibility of each of the other suspects having committed the murder.

The type of instruction that radical constructivists hate the most, it turns out, is the most eff in teaching students. There have been thousands of poorly-done education research studies every so often, some rigorous ones are conducted - program evaluations that use random assignment. With a couple dozen or so random assignment evaluations of thematic educatio programs having been performed, the evidence clearly shows that highly structured, highly s instructional methods work the best. These are programs employing workbooks and lots of repetition, methods that look like "test preparation," methods that look like "drill and kill." constructivist programs show no evidence of success.(2)

basing important decisions about students and schools on a single indicator is not fair

"...a single indicator." "...the use of a single indicator to assess learning or to make decisions tracking, promotion, and graduation violates the ethics of the testing profession." "The scor loom so large that they overshadow discussion of other, more telling indicators of quality of education."

The Rice professor must know that students are not held back because of a single poor performance on the TAAS, they are given several more opportunities to pass the exam. Mo she also knows that the TAAS is a pretty low-level test. It isn't as if students are getting den diplomas because they cannot pass a test based on 12th-grade level material, it's material at lower level of difficulty than that. If a student cannot pass the TAAS, and has a good grade-point-average, either there is something funny about grade point averages at his scho has a severe case of test anxiety.

Scores on a standardized test are not a valid measure of student achievement

"Highly touted rates of improved scores (for example, that Texas was described as in the to "most improved states") mask the fact that even after such "gains," Texas student were still below average, registering lower than 21 of 40 participating states." "The scores loom so la

they overshadow discussion of other, more telling indicators of quality of education, among the degree of segregation, the level of poverty, or the number of student graduating, taking SAT, and going to college."

You just can't win with these guys. Texas is one of the most improved states in the country neutral National Assessment of Educational Progress (NAEP), AFTER implementing the T but our Rice professor says it is still not among the top states. So, she means to imply that the TAAS Texas would have vaulted to the top ranks? ...even though it had rested consiste near the bottom for decades before? I don't get it.

As for her, more "valid," outcome measures...since when did degree of segregation and leve poverty become measures of educational achievement? Moreover, most citizens of Texas ar probably more concerned that their students graduate with skills that will enable them to lea productive lives, or take the SAT with a chance of doing well and going to a good college a making good use of the opportunity. Graduating from high school without having learned h read and write does no one any good, least of all the graduate.

Really, you have to admire the people behind anti-testing "semantic distortion," though. (Pl somebody help me think of a better phrase.) Most of the education "research" that attacks t of standardized tests is more pseudo-science than science, but it is usually very strongly wor uses lots of terminology that sounds sorta scientific, is written by folks with academic crede from (often) prestigious universities, and these folks have developed an expertise in phraseo that works (to persuade naive (or ideologically sympathetic?) journalists, for example).

I would describe what they do as research-by-rhetoric, with rhetoric based on tortured sema It's dishonest, but, to a large degree, it seems to work for propaganda purposes. Their research-like prose is to real analysis what guar gum is to diet food, filler with no nutritional contribution. As they say, you can get data to say anything if you torture them long enough. you can get words to mean anything if you torture them long enough. But, hey, it works.

Don't blame the Rice University professor. She's just trying to get ahead in her career and, a I can tell, she's doing all the right things to make it happen. Several years ago, I read a "benefit-cost analysis" done by an education professor in the 1980s that criticized the popul Texas requirement that their teachers pass (what was, by all accounts, an incredibly easy) ba literacy test. This particular education professor hated that test and felt abhorrence at the au of the people of Texas in requiring their teachers to pass a primary-grade level literacy test. study was awful, horribly biased, representing the most egregious case of lying with statistic ever read.[3]

Guess what has become of that education professor? She has received several prestigious outstanding researcher awards from the professional associations of which she is a member last year, she was elected president of the huge American Educational Research Association was, in essense, named the country's leading researcher on education issues. Perhaps the Ri University professor can expect a similar reward for her efforts.

Richard P. Phelps is the author of *Why Testing Experts Hate Testing* (www.edexcellence.net/library/Phelps.htm)

1. See http://www.rethinkingschools.org/Archives/14_04/tex144.htm and http://www.law.harvard.edu/groups/civilrights/conferences/testing98/drafts/mcneil_valenzu

2. See http://www.edexcellence.net/library/carnine.html and http://www.edexcellence.net/library/bbd/better_by_design.html.

3. See www.edexcellence.net/issuespl/subject/standar/testbash.html.

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 3

The Education Press's Cop-Out on Student Testing

by Richard P. Phelps

It was a typical *Education Week* article on testing, "FairTest Report Questions Reliance on High-Stakes Testing by States" (Jan.28, 1998). It was typical in that it: gave the anti-standardized-testing advocacy group that calls itself the National Center for Fair and Open Testing (FairTest) another headline; treated one of their reports as a serious piece of research; and, for journalistic balance, approached the country's next most prominent anti-standardized-testing advocacy organization, the Center for Research on Evaluation, Standards, and Student Testing (CRESST) for an evaluation of the FairTest report. That evaluation shouldn't have surprised anyone: "...all of the researchers interviewed agreed with FairTest's contention that research evidence supporting the use of high-stakes tests as a means of improving schools is thin."

In their report, FairTest noticed that states with high-stakes minimum-competency test graduation requirements tend to have lower average test scores on the neutral National Assessment of Educational Progress. This was offered as evidence that high-stakes tests *cause* lower achievement. They made no effort, however, to control for other factors that influence academic achievement, and the relationship between cause and effect was just assumed to run in the direction FairTest wants. Honest observers would conclude the direction of cause and effect to be just the opposite - poorly performing states (mostly in the South) initiated high-stakes testing programs in an effort to improve academic achievement while high performing states (mostly in the Midwest and New England) did not feel the need to (until recently, that is).

So, how "thin" is the "research evidence supporting the use of high-stakes tests as a means of improving schools?" The work of the Cornell labor economist John Bishop does not get the attention the education press bestows on FairTest and CRESST. Yet, in a series of solid studies conducted over a decade, Bishop has shown that, when other factors that influence academic achievement are controlled for,

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW
Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health

Texas Reading Institute

No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

students from states, provinces, or countries with medium or high-stakes testing programs score better on neutral, common tests and earn higher salaries after graduation than do their counterparts from states, provinces, or countries with no or low-stakes tests. He ran multiple regressions with cross-sectional data using common, neutral tests such as the International Assessment of Educational Progress, the Third International Mathematics and Science Study, the SAT and the NAEP.

Bishop even studied the very same relationship that FairTest did, only he looked at it in some depth. He and his colleagues used individual-level data from the National Education Longitudinal Study (NELS:88) and High School and Beyond (HSB) databases, controlled for socioeconomic status, grades, and other important factors, and compared the earnings of graduates from "minimum-competency" testing states to those from non-testing states. They found that test-taking students earned an average of 3 to 5 percent more per hour than their counterparts from the non-testing schools, once other factors were controlled for. And the differences were greater for women, with as much as 6 percent higher earnings for those who had taken the tests.

While John Bishop has done the most work in this area, he's just one of many. Jonathan Jacobson did a similar study with minimum competency test states and NELS data and found strong benefits for low-achieving students. Robert Costrell has demonstrated with tight logic and elegant mathematical models how strict standards and tests with stakes improve learning given the relative sets of incentives with and without meaningful tests.

The psychologist and attorney Barbara Lerner has written many stories and reports citing the benefits of high-stakes standardized testing in the states, though one won't find those stories in the education press. Michael Podgursky, Dale Ballou, and Ron Ferguson have written much about the benefits of teacher testing, but they have to write letters to the editor to be heard in the education press. David Murray has done the math on testing opponents' proposals to eliminate the SAT in college admissions (and rely on grades alone) and found that black students would end up worse off, contradicting the anti-testers' claims.

Outside of education, John Hunter and Frank Schmidt are perhaps the best known among the hundreds of personnel psychologists who have, for decades now, conducted thousands (yes, thousands) of controlled studies demonstrating that fairly general achievement or aptitude tests provide employers a more reliable prediction of employers' future performance than any other screening measure commonly used. The education press knows nothing about their work.

Parental Opt-Out Form

Public Policy Org

NEW Publications francaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research
The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

Anyone with basic statistics knowledge can compare states with high-stakes tests against those without during the 1990s on their changes in NAEP math and reading scores over the decade. The difference is strong and dramatic - states with high-stakes tests are improving at a statistically significant, higher rate than their counterpart states. David Grissmer looked at the improvement in the two most heavily tested states in the country, which also happen to be the two most improved states in the country -- Texas and North Carolina -- and, controlling for other factors, found that the high-stakes testing regimes were producing dramatic gains in achievement.

In a long May 31 *EdWeek* article on Texas' testing program, a reporter wrote:

"On the seven NAEP tests given to 4th and 8th graders between 1990 and 1996, Texas and North Carolina made the largest average gains in the nation, according to a *widely reported* 1998 analysis by David W. Grissmer, a senior managing scientist at the RAND Corp., a Santa Monica, Calif.-based think tank." [italics mine]

"Widely reported" everywhere but in the education press, that is. The reporter never talked to Grissmer and *EdWeek* gave no coverage to the report, not even a paragraph in its "Report Roundup" section.

Grissmer has now completed another study of all 50 U.S. states and found high-stakes testing regimes to strongly predict academic improvement on the NAEP. This study has also been "widely reported," for example, in EducationNews.org. Will the education press give it any attention at all? We shall see.

Virtually no attention in 12 years to the research work on testing benefits. Yet, *Education Week* gives abundant coverage to any alleged research fostered by the misnamed FairTest. It runs magazine-length features on the prophets from Cloud Cuckoo Land who urge us to ban all "extrinsic" rewards from our schools. Without "extrinsic" rewards, humans and human civilization wouldn't exist.

I don't charge that *Education Week*, or its counterparts in the education press, are intentionally biased, but something is surely "unbalanced" here. I used the search engine at *Education Week's* web site (which is wonderful, by the way) to do a little arithmetic. I judged who or what I thought were the most prominent anti-testing sources and the top researchers who work on the benefits of testing. Granted, my choices may not be perfect, but judge for yourself, here they are, thirteen in each group: anti-testers: (FairTest, Neill, Schaeffer, CRESST, Linn, Baker, Shepard, McDonnell, Koretz, Smith, Madaus,

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

Haney, Kohn) and testing benefits researchers (Bishop, Jacobson, Costrell, Hunter, Schmidt, Podgursky/Ballou, Ferguson, Lerner, Grissmer, Phelps, Gandal, Cizek, D. Murray). If you don't like my choices, suggest some different names.

I ran *EdWeek*'s search engine on every name and counted the number of "hits" (i.e., pieces in *Education Week* issues containing the person's name), going back to 1988. This isn't an exact science, either in the search method or in the character of what is retrieved. I did make sure that I didn't double count hits, however, and I did try to learn if an individual sometimes was referred to with a middle initial and sometimes without. While all hits contained the person's name, however, not all those pieces were necessarily about testing.

So, this method is not a perfect measure of *Education Week's* "balance" on the two sides of the standardized testing debate, and I intend to continue to refine it. Nonetheless, this little research effort is revealing. Here's the score: anti-testers, 430; benefits researchers, 92. Five of the benefits researchers had no hits at all; most had less than five.

One might argue that perhaps the anti-testers make more of an effort to be heard. Well, not really. The hits for the benefits researchers are dominated by letters and op-ed pieces sent into *Education Week*, while the hits for the anti-testers are full of solicited quotes. The benefits researchers have to push hard to get heard; the anti-testers get convenient phone calls from *EdWeek* reporters. Of the thirteen anti-testers, twelve of them had been called at least once by *EdWeek* reporters; some had been called dozens of times. Of the thirteen benefits researchers, only one - Matt Gandal -- has ever received a phone call from *EdWeek*.

Thank goodness for the intrepid Matt Gandal, formerly of the American Federation of Teachers and now of ACHIEVE, who alone has been selected by the education press to present to the public the point of view that standardized tests may not be evil incarnate and that, by gosh, they may actually have some benefits.

Indeed, if it weren't for the Commentary Editor at *EdWeek*, Sandy Reeves, who selects the OpEd pieces and letters to the editor, the testing benefits researchers would not have had any opportunity to express themselves at all, for over 12 years.

While one common type of testing story in the education press uses one anti-testing group to "balance" a report on another anti-testing group, as described above, another common type of media story on testing gets started by an attack on a test. The resulting published story then focuses on the attack and the defender, speaking for the

"other side", is usually a local or state test administrator, just trying to defend the job she's doing. The attacker is usually given a lot of space, while the defender is given little. The attacker usually makes academic arguments and cites "the research." The defender says their tests are well-done and responsible. The "research" claims are never countered.

I telephoned several education press reporters to try to understand why testing stories get set up this way. They replied that they do not know of any advocacy group on the other side of the testing issue that can balance the point of view of the established (and well funded) anti-testing groups, such as FairTest and CRESST. They added that these groups are also very reliable: they keep up with the issues and they return phone calls promptly. As Greg Cizek has written: "the measurement profession has made no corresponding, popular, accessible, public defense of its mission or of testing." If the reasonable voices never speak up, what is the public to think? But, how can they speak up if they never get a phone call?

The journalists are doing what is *convenient* for them and, in the process, they are letting the side with the money and the organization determine the debate. CRESST receives millions of dollars a year from the U.S. Education Department. If you count the National Research Council's Board on Testing and Assessment, which might as well be considered a CRESST subsidiary given its common personnel, its wholesale (and virtually exclusive) reliance on CRESST "research," and its common benefactor, there's another million dollars a year at a minimum. The anti-testers at Boston College's CSTEEP just won a $1 million grant from the dependably naive Ford Foundation. Fairtest gets a half million a year from The Ford and Joyce Foundations alone.

This huge treasure trove funds the "research" that these organizations produce, as well as the infrastructure for coddling journalists. The education press can call FairTest, CSTEEP, and CRESST and get their calls returned right away because these organizations pay for full-time staffers to handle press relations.

What are the resources available to the other side - those of us who do research on the benefits of testing? Well, perhaps you can count Matt Gandal's salary at ACHIEVE and the staff overhead that's dedicated to his work -- at most $100,000/year

Good reporting is surely about working hard, writing well, and establishing contacts but, just as surely, it must also be about knowing the agenda of every source. Without that knowledge, stories can be lopsided. On issues where all sides have equal amounts of money and organization, it might not matter. But, on issues where the sources on one side have money and organization behind them,

and the sources on the other side do not, it could matter.

One such lopsided issue area is education research. Research on education is not like research in other public policy issue areas. For one thing, much of it is conducted by graduates and professors of education schools, who have a vested interest in maintaining the current structure of U.S. education. On many education policy issues, this vested interest does not curtail the discussion, because one can still hear more than one side to the debate within the education research establishment itself. But, on two issues in particular - school choice and "high-stakes" student testing - the vested interest determines the debate. What is different about these two issues is that they threaten the very existence of the education research establishment.

Much of the research on testing is advocacy presented as "technical" research from "independent" sources, that is neither technical nor independent. The dirty little secret of the education research establishment is its profound self-interest in the issue and its ideologies of radical constructivism and radical egalitarianism.

The self-interest is easy enough to understand. Without standardized testing, no one can know how our schools are performing. So long as no one knows, few will protest. So long as no one protests, education professors will be held to no standards in training teachers. Held to no standards, education professors can do as they please - teach ideology rather than pedagogy or subject matter or, perhaps, teach as little as they please. Nice work if you can get it.

The ideology is easy to understand, too, if you think about it. Education professors are not randomly picked into their livelihood, they are self selected.

Reporters should not expect objective comments on the benefits of standardized testing from CRESST or FairTest anymore than one would expect objective comments on the dangers of a carcinogen from an industry-funded research lab for the industry that produces that carcinogen. Most people, in the end, defend their self-interests, either because they sincerely believe their interests are good for everyone else, or because they worry about keeping their jobs and getting their kids through college.

Nonetheless, there are some courageous education professors who would like to speak out against the education research establishment, but are afraid to. Afraid of having their papers rejected at journals. Afraid of being denied tenure. Afraid of being denied promotions. Afraid of being labeled elitist or racist. Afraid of being compared to the authors of *The Bell Curve* or to the racists in the 1920s and 1930s

who wished to misuse standardized tests to uphold their perverted theories. Afraid of the inquisition. The education press could talk to them anonymously, but they don't.

Moreover, not all testing experts are education professors. There are in fact hundreds of qualified testing experts working for national, state, or local agencies (not to mention the experts working for organizations that develop tests *under contract* to governments). But, they are contractually and ethically restricted from expressing their views regarding testing policy. The education press could talk to them anonymously, but they don't.

In summary, I believe that *Education Week* is biased, one could say, *in its procedures*. (And, if anything, its competitors are worse.) I also believe that the damage caused by this bias is enormous. Just in the past year, I have read many awful, one-sided pieces of journalism on testing that cited "the research" on high-stakes tests and wondered where these people get their weird ideas. I then noticed that many journalists who do not normally work on education stories get their background information on testing from the education press. They trust it. Indeed, one can find direct links to *Education Week's* Web site in the education sections of other media Web sites.

I think that I am not going out on a limb by asserting that most journalists who have worked on education testing stories this past year believe the following: multiple choice tests primarily elicit factual recall; time spent preparing for or studying for exams is not "real learning"; what happens inside the classroom when there are no high-stakes tests "corrupting" "natural" instructional practice is necessarily superior to what happens under the pressure of high-stakes tests; drills, practice, worksheets, and teacher-centered, highly-structured lessons are bad instructional practice; and politicians and businesspersons are the prime movers behind high-stakes tests, whereas most teachers, parents, and students oppose them. Indeed, if my only source of information on the subject of testing was the education press, I would believe all of this.

All of the statements above, however, are nonsense. They are not just not truthful, they are the opposite of the truth, as can be easily demonstrated with common sense, easily produced evidence, or solid research of the type the education press generally ignores. What does *Education Week* want its legacy on testing to be? ...that it helped a bunch of self-interested or ideologically motivated charlatans deceive the good citizens of this country to act against their own best interests and the common good? That's to be the legacy of "education's newspaper of record?"

Post script:

Aug.2 Education Week this week has run a story on David Grissmer's latest study. How did they handle the finding that states with high-stakes standardized tests show more improvement on the NAEP? They completely ignored it and focused on other findings.

Richard P. Phelps is the author of *Why Testing Experts Hate Testing* (www.edexcellence.net/library/Phelps.htm)

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 4

Appearing *Exclusively* every Tuesday

That "Backlash" That Testing Opponents So Desperately Crave

Richard P Phelps

Did you know that there is a public "backlash" against academic standards and high-stakes tests? You're supposed to. Some of the regular crew of testing opponents and some sympathetic journalists insist that it's true. It's been pronounced to exist by one testing opponent in the (dependably naive) *Atlantic Monthly*, so it must be true. The *Atlantic Monthly* wouldn't just make something up.

It's also been declared to exist by the director of the Board on Testing and Assessment of the National Research Council, a group that extols anti-testing research no matter how poorly done, and discounts (or ignores) contrary research no matter how well done. The director is paid well by us taxpayers, but his loyalties are fully directed toward the vested interests on testing issues.

Let's examine their evidence for this backlash, conveniently provided in a list by Peter Schrag in his *Atlantic Monthly* article.

First, 300 students in that bellwether state of Massachusetts sat out a required state examination. So, 300, out of 18 million high school students...that's a proportion of 0.0000167 of our nation's high school students, or 0.00167 percent.

Second, Wisconsin legislators refused to fund an exit examination that had been approved two years earlier (by a different legislature, which may have had a different composition and preferences). A compromise was reached in which student graduation will be determined on a variety of criteria, not just the exam. But, the exam stays, and Wisconsin has never before had an exit exam.

Third, in Virginia, Republican legislators proposed (and the governor accepted) to let students graduate from high school if they can pass some other exams as well as Virginia's new high-stakes exams. This is interpreted as backtracking. What are the other exams? the International Baccalaureate, Advanced Placement exams, and SAT Subject-Area exams, all more difficult than the Virginia exams. Moreover, even students opting for these other exams in some subject areas must still pass a Virginia test core for the primary subject areas.

Oh, and there's a "grassroots" group of the public opposed to the Virginia exams, too. As it turns out, the person who appears to be that group actually works for FairTest, which is

Ed. Links

Ed. Org

ELA

Grammar

Hispanic Links and Resources

History

Home School

Job Postings

K-12 Science

NEW Learning Disabilities and High-Stakes Standardized Tests

Math

Media Links

Mental Health

Texas Reading Institute

No Child Left Behind

National Reading Panel Report

Phonemic Awareness

Reading Kits

Reading Tests

Ideas Practicas Para Padres de Familia pdf file

Special Education Resources

TAD for English/Language Arts/Reading

LA Times Special Report Special Education a Failure on Many Fronts

about as grassrootsy an organization as a K Street public relations firm. Several months ago, another group of citizens, annoyed that the small group of complainers was getting so much attention from the press, formed their own grassroots group of supporters of Virginia's tests. Perhaps they represent a "backlash backlash."

Fourth, three former members of the State Board of Regents complained that the new graduation policy was too rigorous, reported the *New York Times*. A few months ago, the same paper reported that several members of the current board were complaining that the graduation test was too easy and the passing rates would be too high.

Fifth, in Ohio, some people opposed to state tests are circulating a petition. *Wow*.

The rest of Schrag's article is much the same. A bill was introduced in this state, some people in another state said something. The testing programs in some states have been influenced by compromise (like virtually every other piece of legislation ever considered in every free legislature in every jurisdiction throughout history). And, don't forget that those mainstream populists, Alfie Kohn, Gerald Bracey, and the FairTest crew, are opposed to high-stakes testing, too. That's our movement.

Schrag's article is also filled with the usual testing opponent contradictions: the tests are too hard and unfair to the slow students/the tests are too easy and not worthy of high standards; it's not fair to judge students based on one single test/with no mention that high-stakes states generally give students several to many chances to pass a fairly low-level test; that parents will be opposed to these tests when they affect their own kids/with no mention of the fact that the parents have been asked this question and a large majority stand resolutely in favor of high-stakes anyway; it's the general public that's opposed to tests/the largest organized meeting of testing opponents took place at Teachers' College, Columbia University.

So, some well-known opponents of high-stakes standardized testing tell us "a backlash has begun." And, some journalists believe this? Heck, I'm a really great guy, and I've had a harder life than I might have liked, so people should send me money. How about picking up that story?

For the most part, the alleged "backlash" is made up of the same people who have always been against standardized testing - some misguided idealists and, more prominently, those vested in the current system of public education who resist change that would be inconvenient for them personally. Sure, one can find a scattering of citizens and politicians opposed to standardized testing who do not fit into those two categories, but you can find 5 percent of the population opposed to any given issue. It's not likely that the alleged "grassroots" movement against testing can claim even that proportion.

It is clear that testing opponents want you to think there's a backlash. They know it won't work to tell you the true reasons why they are opposed to testing; their opposition has always been couched in propositions of defending the defenseless who they claim are harmed by tests. As such, it's always been inconvenient to them that the vast majority of Americans strongly support high-stakes testing. As Peter Schrag himself admits:

"The movement has a long way from achieving critical mass. The two most prominent lawsuits brought to date...have failed. The boycotts are still small, and polls, by Public Agenda and other organizations, continue to show that 72 percent of Americans - and 79

Parental Opt-Out Form

Public Policy Org

New Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

Agenda and other organizations, continue to show that 72 percent of Americans - and 79 percent of parents - support tougher academic standards and oppose social promotion 'even if [the outcome is] that significantly more students would be held back.'"

For the sake of completeness, students and teachers should be added to the list of groups overwhelmingly in favor of high-stakes standardized testing.

If testing opponents can't really have the public on their side, then, they have to create an imaginary public that's on their side, or about to be on their side. The public against testing isn't large now, but, darn it, testing opponents are certain that oppostion is growing.

As evidence of this change in the public's thinking, a few months ago, the American Association of School Administrators paid a "hired gun" polling firm to do a "quick and dirty" poll. To much fanfare, the AASA announced newfound, large-scale public opposition to high-stakes standardized tests. Here's what I wrote about that poll at the time:

"Collectively, high-stakes standardized testing's opponents comprise a small part of the citizenry, but some of them are rather outspoken and manage to get some journalists' attention. These testing opponents tend to be of three affiliations: principals and local school administrators; education school professors; and those who dislike tests for ideological reasons. I used to feel some sympathy for the principals and local administrators, since they often are in the front line of blame when students test poorly and, as they argue themselves, they do not maintain complete control over how students test. My sympathy waned when I saw Paul Houston's piece in the *Times*.

"Why did the AASA poll get such different results for the general public's attitude toward high-stakes standardized tests [from the several dozen other polls conducted on the issue]? Two possibilities: the attitude of the general public has changed drastically in recent months and the poll was done poorly or, worse, was rigged.

"Starting with the latter possibility first, I telephoned the AASA's polling firm (Luntz/Lazlo) requesting their methodology report and a copy of the survey instrument. I've made this kind of call dozens of times. I was refused. I've never before encountered a survey firm that did not have a methodology report available and that refused to provide such essential information. I asked directly for their survey response rate. Ms. Lazlo did not know what it was, but offered that their firm usually called four times as many persons as complete their interviews.

"So, *maybe*, they achieved a 25 percent response rate of registered voters, their target group - 1 out of every 4 voters in their random sample responded. The U.S. Education Department requires a 90 percent response rate in its surveys. The chief survey methodologist at the U.S. General Accounting Office, the country's chief program evaluator, once told me that, with a response rate of 70 percent, a surveyor is "just starting to get respectable." Surveys with low response rates may only represent that segment of the population with an ax to grind. Virtually all the polls I reviewed in my study of testing attitudes achieved response rates of at least 50 percent and, in most cases, over 70 percent. (The only ones that didn't were two done by another vociferous anti-testing organization, the *Phi Delta Kappan* (PDK) magazine (their own survey of teachers, not the one of the public they hire Gallup to run), with response rates of around

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

20 percent.

"But, could public opinion on high-stakes standardized tests have changed dramatically just recently? After all, new testing programs in some of our larger states are just now starting to bite (i.e., hold kids back if they fail). If the AASA survey conducted by Luntz/Lazlo were the only one available from recent months, that could still be an open question. It's not, and the other recent polls achieved much different results."

Incidentally, the pundit Ariana Huffington has urged voters to not participate in these quick and dirty polls with the wretchedly low response rates. She argues that the lonely and unemployed (those most likely to be at home to answer the phone) are representing all of us.

Education Week is helping in its way to promote the story, too, declaring a backlash to be a fact, interviewing only testing opponents for views on the issue, and turning this one bogus poll by one self-interested group into "several" polls showing definite public opposition to high-stakes testing.

I think the following story is more plausible than the "backlash" movement: As more states implement their tests, those less-than-5-percent of the public opposed to tests, plus the usual coterie of education professors and ideologues who have always opposed accountability measures, appear along with the tests themselves. After all, they do exist in every state. There's no increase in the proportion of testing opponents nationwide, it's simply the same proportion we've always had, but it pops up with each new exam as journalists in every state try to cover "both sides" of the story. Testing opponents want to call this growth, or a movement. They don't tell you that, in parallel, state after state is actually successfully implementing the high-stakes exams the testing opponents dread.

If anything, isn't it really more likely that what little public opposition there is will diminish over time, as lawmakers tweak and fine tune their testing programs in response to their initial experiences with the testing programs? Given that testing programs are complicated and difficult to implement and given that (unlike in the rest of the industrialized world where high-stakes testing has been the norm for decades) they are still new in the United States, it's remarkable that there haven't been more problems and more complaints. The fact that there has been so little testifies to the strong public support for standards and high-stakes. Journalists could draw attention to this, the huge, steadfast majority, but some would rather characterize the exception as the rule.

As Deborah Wadsworth of Public Agenda (an organization that conducts professionally responsible polls (with high response rates, and valid and reliable questions)) said recently at a conference:

"...public opinion endorses higher academic standards. 'Support for raising standards is rock-solid in every part of the country and among people of every background,' she said. 'Members of the press insist on finding a backlash to standards that we do not find.'"

Richard P. Phelps is the author of "Why Testing Experts Hate Testing" (http://www.edexcellence.net/library/phelps.htm)

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

Test Bashing, Part 5
More on Texas Testing and School Reform
Richard P. Phelps

I'm attaching an article written by Jay P. Greene, a scholar who has worked often with Paul Peterson, taught at the University of Houston, and is now affiliated with the Manhattan Institute. His research is solid, his focus keen, and his writing clear; what he writes is worth reading. This article appeared in the Manhattan Institute's *City Journal*, Summer, 2000 (vol.10, no.3)http://www.city-journal.org/html/10_3_the_texas_school.html

Just one comment: I'm not familiar with the incidents to which Greene refers in his last two paragraphs, where he criticizes teachers' attitudes toward testing. Personally, I believe that teachers in general and the American Federation of Teachers in particular have the most important forces in our nation in the effort to implement high-stakes student testing programs. Testing opponents often claim teachers as their allies: surely some of them are, but most of them are not.

The Texas School Miracle is for Real

Richard P. Phelps is the author of *Why Testing Experts Hate Testing*

http://www.edexcellence.net/library/phelps.htm

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

Shop Here for Books
amazon.com

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 6

The Education Press's Cop-Out on Testing (continued)

Richard P. Phelps

The last time I wrote about the education press I received several Emails informing me that journalists really are objective and, besides, the newspaper I focused on then, *Education Week*, is highly unrepresentative of education reporters' work as a whole. I may, indeed, have been misleading in my commentary because, by "education press," I was referring to those venues that focus on education issues exclusively. Nonetheless, I got the point and decided that I would look at the education press as a whole. As I do not have the resources to search the archives of every newspaper in America, I chose to look at the website of an organization that claims to represent all education journalists, that of the Education Writers' Association (EWA). Here's what I found:

In the "Hot Topics" section is a subsection on testing that contains 26 paragraphs. About half are factual or neutral to any testing debate, but the rest are not. Those paragraphs feature quotes from commentators and presentations of opinion. Ten paragraphs are devoted to the anti-testing point of view, while just two offer a pro-testing viewpoint. Representing the anti-testing point of view are the usual coterie of advocates and professors: the National Center for Fair and Open Testing (FairTest), Alfie Kohn, Peter Sacks, Deborah Meier, the National Center for Restructuring Education, two Latino civil rights groups, and professors Robert Linn and Robert Hauser. Representing an opposing viewpoint are the American Federation of Teachers (AFT) and me, of all people.

In the same subsection is a page devoted to an alleged anti-testing "backlash" organized by independent local groups of students and parents, which the EWA accepts as a fact, rather than as a ruse mostly promoted by the same old anti-testing organizations ("grassroots organizations are rising up against tests"). You can, however, find most of these "independent grassroots organizations" on FairTest's web site, where you'll also notice that most of the leaders of these "independent grassroots organizations" work for FairTest. Just compile a list of the leaders of all these "independent grassroots organizations" and then look at a list of FairTest's state coordinators. You can also find some of the documents that FairTest and these "independent grassroots organizations" use to harass state and local testing directors and misinform the public. The documents look a lot alike, like maybe they haven't "risen up" so "independently" after all.

In the "New Research" section of the EWA's website are listed two sources, a Harvard Graduate School of Education conference transcript and a primer on testing written by Gerald Bracey for the American Youth Policy Forum. The transcript features the comments of three staunch opponents of high-stakes testing (Angela Valenzuela, Ted

Ed. Links

Ed. Org

ELA

Grammar

Hispanic Links and Resources

History

Home School

Job Postings

K-12 Science

NEW Learning Disabilities and High-Stakes Standardized Tests

Math

Media Links

Mental Health

Texas Reading Institute

No Child Left Behind

National Reading Panel Report

Phonemic Awareness

Reading Kits

Reading Tests

Ideas Practicas Para Padres de Familia pdf file

Special Education Resources

TAD for English/Language Arts/Reading

LA Times Special Report Special Education a Failure on Many Fronts

Sizer, and Linda Nathan) and one journalist who could, arguably, be called neutral. For his part, no argument is possible that could portray Gerald Bracey as neutral. He is perhaps, the country's most vociferous defender of the status quo and a master at using statistics selectively to support his points. He calls those he disagrees with names, annually awards them "rotten apples" in his *Phi Delta Kappan* column, and labels many of them "right wingers." I think it is in very poor taste that the American Youth Policy Forum chose him, of all people, to write their primer on testing. His primer is not as slanted as most of what he writes, but it is definitely still slanted.

So, there you have it. The EWA web site gives prime time to pretty much the whole range of anti-testing opponents. Between the "Hot Topics" and "New Research" sections, 13 anti-testing opponents are featured. How many commentators are represented who would argue that testing may not be so horrible and terrible and, by gosh, may actually have some benefits. Just two of us, the AFT and me...and even the paragraph about the report I wrote is misleading; they got both of the arguments they cite wrong.

Our country boasts the most knowledgeable psychometricians and the most advanced psychometrics in the world. Other countries look to the U.S. for expertise and are starting to adopt the advanced technical methods developed here (yes, multiple choice tests). Do U.S. journalists talk to these, the world's most knowledgeable testing experts? No.

Granted, there are many testing experts, including those who dislike testing as well as those who do. The latter outnumber the former by a large margin, however. The thousands of testing experts who like testing are out developing and administering tests, and journalists don't talk to them much. The relatively much smaller number of testing experts who hate testing stay behind in academe, and journalists seem to talk to them a lot.

State and local testing directors tend to get little attention from the press, unless it's to defend against a false charge from anti-testing "experts." But, testing directors are the most fair-minded people I know. They also tend to be extremely well-educated and technically proficient scientists. They chose their profession because they believe that more information is good and that standardized tests provide useful information that cannot be obtained by any other means. They are also obsessively concerned with fairness. They believe, as most citizens do, that an evaluation of a student or a school will be better with the addition of information from standardized tests. Moreover, generally they are politically liberal. I would guess that you would have a hard time finding many Republicans among them, much less "right wingers."

Yet, they are continually harassed by the organization many journalists seem to like, the misnamed National Center for Fair and Open Testing (FairTest). They are portrayed as mean ogres who wish to punish kids, and stooges of the corporate establishment that testing opponents charge is the real impetus behind high-stakes tests.

Which begs the question - how do those who stubbornly resist any change in the status quo merit the label "liberal", while those who press for change, a more open system with more useful information, and more power for students, parents, and teachers, merit the label "right wing" or "conservative?" We all know about the studies that purport to show that a much larger proportion of the journalist corps considers itself "liberal" than does that of the general public. Is that why journalists are attracted to testing opponents, because they declare themselves to be "liberals" fighting "right wingers?"

Parental Opt-Out Form

Public Policy Org

NEW Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

because they declare themselves to be "liberals" fighting "right wingers?"

With any other public or private monopoly, liberal journalists would advocate monitoring, evaluation, full disclosure, and regulation; that's certainly true with kilowatts of electricity and cubic feet of water. But, with our own children, where one would logically think monitoring, evaluation, full information, and regulation are even more important, many journalists advocate leaving the system untended and fully within the control of the interests who profit from the status quo.

Without standardized tests, administered by an authority higher than and outside the school district, we have no way of knowing what is going on inside the schools. And that is how status quo defenders want it. If we don't know, who's to complain? And if no one complains, the vested interests can do as they please with our children, or do nothing at all.

The kids who suffer the most under status quo intransigence are the poor. They are left alone to drift in unstructured courses and unstructured schools, sitting passively for 12 years (if they can stand it for that long) while the defenders of the status quo collect their paychecks. The poor kids are socially promoted year after year and told they are doing just fine, until they eventually graduate without being able to read or write. Where can these students get the structure and help they need to actually learn something useful? In states with high-stakes tests, that's where.

In states without high-stakes tests, standards don't matter, meeting standards doesn't matter, and the poor kids get forgotten, and we get told by those in control that all is well with the schools.

To be completely thorough, however, I must admit that the EWA website includes links and references to other organizations where one might find more information about testing and, to be fair, some organizations that do not hate testing are listed. But, the majority of those listed are testing opponents.

If I had my way, I would add links to the following organizations and publications to the EWA's website. They are listed in order of my perception of their importance. They are either neutral in the testing debate or advocates for testing.

Achieve

National Council for Measurement in Education (NCME)

journal: *Educational Measurement: Issues and Practice* (EM:IP)

National Association of Testing Directors (NATD)

any test developer, of which there are many

National Governors' Association (NGA)

State School Boards Association (SSBA)

Education Leaders' Council (ELC)

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

journal: *Educational and Psychological Measurement* (EPM)

American Psychological Association (APA)

Core Knowledge Foundation

Division H of the American Educational Research Assn., and none of the rest of AERA

National Assessment Governing Board (NAGB)

Human Resources Research Organization (HumRRO)

Society of Industrial-Organizational Psychologists (SIOP)

journal: *The Industrial-Organizational Psychologist* (TIP)

Brookings Institution

Heartland Institute

Hudson Institute and The Modern Red Schoolhouse

Manhattan Institute

No Excuses

StandardsWork

Mathematically Correct

Pioneer Institute

Note that none of the organizations listed above belong to the alleged corporate conspiracy to impose high-stakes tests on defenseless children in order to punish them and destroy the public school system.

Within the education establishment itself the organizations opposed to high-stakes student testing tend to be those that represent school- or district-level administrators and education school faculty. Those in favor of high-stakes student testing are state-level administrators and teachers (of course, teachers often have different feelings about testing programs that put them on the line for their students' performance, over which they do not have complete control). Outside of the establishment, the vast majority of the general public, parents, and even students favor high-stakes testing. Who outside the establishment opposes it? Maybe journalists.

If a journalist wants to write an article attacking the use of high-stakes standardized tests, here are the organizations to talk to that have not already been mentioned above: those representing school or district administrators, such as the American Association of School Administrators; those representing principals, such as the National Association of

Elementary School Principals; those dominated by district administrators, such as the National School Boards Association and the Parent-Teacher Association; the American Society for Curriculum Development (counselors and media folks); Phi Delta Kappa; the American Educational Research Organization, the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA, and the National Research Council's Board on Testing and Assessment, which is virtually a subsidiary; and the Center for the Study of Testing, Evaluation, and Education Policy (CSTEEP) at Boston College.

Incidentally, a group of several testing opponents at Boston College rather immodestly calls itself the National Board on Educational Testing and Public Policy. I'm thinking of getting together with a few friends and calling myself chairperson of the World Council on Testing, or some such. I'm open to suggestions for other names.

Finally, what have I learned about the education press from my tour of the EWA's web site and my reading of about a hundred newspaper and magazine articles on testing in the past few months? Here's what I think I have learned:

journalists don't read (academic) journals;

journalists generally believe what other journalists say and write;

journalists are hooked on getting their information from advocates and advocacy groups;

journalists generally do not recognize who the vested interests are in the testing debate and accept most of them as neutral and objective sources;

a "grassroots movement" can consist of an infinitesimally small number of people; and

journalists accept most of what testing opponents tell them at face value, particularly "the research" that testing opponents claim supports their cause.

Yes, judging from EWA's web site, I'd say it is biased and unbalanced in its coverage of testing issues. If EWA's coverage is representative of all journalists', I'd say that our country has little hope for objective or balanced coverage of testing issues in the press.

As one final testimony to EWA's alleged bias, I note who supports it. FairTest, the most virulent anti-testing organization in the country, and the largest supplier of misinformation on testing, only lists or provides links to other anti testing groups or individuals. There is absolutely no effort at "balance" in its literature or on its website; it gives its members a consistent diet of just one (very extreme) side of the story. Who's listed among those organizations that FairTest recommends for further information? A bunch of other well-known anti-testing organizations, and, you guessed it, the Education Writers Association.

Richard P. Phelps is the author of *Why Testing Experts Hate Testing*

http://www.edexcellence.net/library/phelps.htm

Post your comments on this article

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing
Test Bashing Series
Intellectual Property and Digital Information in Higher Education
LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us
Advertising
Archives
DailyNews
Past Five Days
Book Reviews
College Aid
Book Reviews
College Aid

Test Bashing, Part 7

The Unfortunate Bias of the Education Writers' Association (continued)

Richard P. Phelps

(Sept. 5, 2000) For those of you just joining us, let me bring you up to speed. Last week, I wrote a commentary criticizing the Education Writers' Association (EWA) as "biased and unbalanced in its coverage of testing issues." The following day, Lisa Walker, the Executive Director of EWA, took issue with my comments. Her commentary can be found 20 or so entries down the list of commentaries, and mine just a little further down.

The impetus for my commentary was my visit to the EWA's website, which I saw recommended at FairTest's website. I wanted to see why FairTest, a rather discriminating evaluator of sources on testing information, was recommending the EWA website. I think I discovered why. I attempted to find anything at the EWA website referring specifically to the issue of testing. My intention was to model the actions of any reporter who might attempt to use the EWA's website as a source for material (and sources) on a story about testing.

Ms. Walker thinks I made a mess of the effort and that I ended up judging her organization unfairly. Below, I list what I believe to be all her assertions, and provide my response to each. I also include two comments made by an EWA member, Linda Loupe, to this website's Bulletin Board, just because I liked them (she also disagreed with a couple of my arguments).

One of Ms. Walker's accusations is that I did not look thoroughly enough at the EWA's website. Though I did look at most of the sections Ms. Walker says I didn't, I notice now that I did overlook a rather revealing box aligned to the left of the main text section on testing (in the "Hot Topics" section of the website). The box bears a large, bold title "Web Sources" and links to five organizations are included. Any reporter only looking at this, by far the longest text section devoted to the testing issue, would be presented only with these links as source references.

Among the five are two unbiased organizations - the Education Commission of the States (ECS) and Catalyst for Chicago Education Reform -- for whom testing is just one of many topics they consider. Both websites are extremely informative, intelligent, and models of balanced coverage. You are just as likely to find entries from opponents on any given issue as from proponents. Further, I think, as do many others, that ECS staffers, who are charged with informing politicians of both political parties and a wide range of political persuasions, are themselves among the best-informed observers of the education scene in the United States. They know all sides of the issues on which they specialize and they keep

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report
Special Education a Failure on Many Fronts

up with them on a daily basis.

The remaining three links in the box are for websites of well-financed, high-profile organizations that concern themselves only with testing. They also happen to be the three most prominent anti-testing advocacy groups in the country (FairTest, CRESST (UCLA), and CSTEEP (Boston College)). Any reporter looking only at the main text section on testing at EWA's website would have these three sources recommended as contacts. In my opinion, they are three least objective organizations on the topic of testing of the many available in the United States. The country's three most extreme anti-testing organizations, zero pro-testing organizations - that's not balanced coverage.

Now, on to Ms. Walker's charges.

Phelps looked at only a small part of the website.

"[Phelps'] conclusion...is simply preposterous, apparently based on a cursory look at two sections of the EWA website and nothing else." "Phelps judged...EWA's website based on one section..."

In fact, I looked at anything I could find at their website that referred to testing. I actually spent quite a bit of time at their website I was so incredulous at the slant it took on testing.

"EWA's work always includes a full spectrum of viewpoints."

Personally, I don't think so. Not this summer. Not on testing. Ms. Walker's statement, however, may reveal part of the problem. Perhaps she doesn't know what the full spectrum of viewpoints on testing is but thinks she does. The EWA website concentrates on the views of advocates who are easy to find, those with the money and organization to promote their views, an advantage almost 100% on the side of the vested interests in the testing debate. Those sources it takes a little effort to find simply are not mentioned at the EWA website.

"...Phelps reviewed a short background paper on testing and an old research section..."

I assume that the "old research section" she is referring to is the one EWA has labeled "New Research." This section links one to two reports, each produced by very prominent opponents of testing. One is a "primer" on testing that is duly slanted to present testing opponents' view on the subject. This is what any reporter who relied on the EWA web site would look at if she wanted to educate herself on the topic.

"...he overlooked an overview on accountability, a sources' section, a second backgrounder on standards, materials from the EWA listserv on testing and standards and reporters' stories - all in the same section."

I looked at anything that looked like it was about testing. I wasn't interested in looking at all the other various topics on education at EWA's website. I was interested in the issue of testing, and testing is the only issue to which I referred in my commentary.

Nonetheless, because it was linked, I did look at a section titled "May 2000 - Standards and Accountability," which is a list of links, to which I referred in my commentary. I also looked at another list of links in a section titled "Education Resources," to which I was

Parental Opt-Out Form

Public Policy Org

NEW Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

looked at another list of links in a section titled "Education Resources," to which I was also referring in my commentary when I mentioned the website's links listings.

I also looked at the "Reporters' Stories" section and, as I look at it again now, I don't see how it is supposed to have reassured me - there are several fairly ordinary reporters' stories there, none of which espouses any particular point of view (see postscript). My commentary, though, was about what Ms. Walker's organization does at its web site, and her organization doesn't write those stories.

There's also a "testing information" link to an ERIC site which does not contain the information claimed for it, and the "Listserves" link takes one to an empty site under construction.

Even now, I cannot find a "second backgrounder on standards," but, in my original search, I looked at material on standards if it was part of a testing section or linked to a testing section. Otherwise, I wouldn't have been interested. For all I know, EWA's coverage of the standards issue represents the epitome of balance, fair play, and thoroughness. That would be wonderful. My issue is with their coverage of testing.

EWA provides fair coverage

"The testing piece he reviewed...is not one-sided. It is a fair statement of issues..."

She and I disagree. I think it is completely one-sided, but anyone else is welcome to read it and judge for themselves. I also believe that it is only marginally a statement of issues.

[She's referring, I assume, to the several paragraphs at her web site with the subheading "Backlash" that, first, declare it to be a fact that there is a new, spontaneous backlash against high-stakes testing and, second, portray the activities of several well-known advocates who have opposed testing for years as evidence.]

"...they [Robert Hauser and Robert Linn] are well-respected researchers..."

Indeed, some testing opponents consider them saints. I don't think a good reporter, however, takes everything he hears at face value. Of course, they, like everyone else, say they are in favor of testing, but only if it is done right, they say. Even FairTest says this. With FairTest it's just an outright lie; they oppose all standardized testing. Robert Hauser's taxpayer-financed National Research Council (NRC) report takes an even more disingenuous tack. It proposes so many restrictions and regulations on high-stakes testing as to make any high-stakes testing infeasible. Just one of Hauser's many proposed regulations is to establish a national board with the power to stop any high-stakes testing at lower levels of government that it doesn't like. This board would be staffed by a particular group of people well known as among the most virulent testing opponents in the country who, by themselves, would stop any high-stakes test. Anyone who thinks Hauser's report is a moderate treatment of the issue hasn't added up the details. And anyone who thinks Hauser's report is a balanced document hasn't noticed who he relied on for source material, and who he ignored.

You would end up in the same suffocating *cul-de-sac* for high-stakes tests if you followed the advice of the over 500 taxpayer-financed reports issued by Robert Linn's Center for Research on Evaluation, Standards, and Testing (CRESST). Their reports and NRC

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

reports bear many characteristics in common (as should not be surprising given their common personnel) - relying on sources they agree with; ignoring evidence inconvenient to their bias; hundreds of efforts to find problems with tests (both real and imagined); not a single effort in over 500 studies to evaluate any benefits of tests (nor to evaluate the problems with the alternatives to tests that they would have us rely on exclusively), and totally ignoring the rest of the industrialized world, where, in most countries, popular high-stakes testing regimes have existed for decades.

Ultimately, however, it doesn't matter if the two had the souls of Albert Schweitzer and Mother Theresa (or that Ms. Walker personally seems to find their views reasonable). They oppose high-stakes testing. They are among 13 testing opponents featured, recommended, and referred at EWA's website, alongside extremely briefs mentions of only 2 proponents. That's not balanced coverage.

"...they [Robert Hauser and Robert Linn] are hardly opponents of all testing..."

She's right; they're not. I confess, I just get tired of typing "high-stakes testing" and often just type "testing" when I really mean "high-stakes testing." The only real debate, however, and the only real policy issue, is about the stakes. (I know, some people harp on test response formats as if the future of human civilization will be determined by whether we choose multiple-choice or open-ended response formats. Indeed, civilization may choose to voluntarily expire over the silliness of that debate). Everybody, but the extreme of the extreme (e.g., FairTest) thinks individual student, no-stakes diagnostic and monitoring tests are OK. And, besides, no government is plausibly going to attain the power to prevent schools from using most no-stakes tests. So, they are not an issue. High stakes is the issue.

"The same section contains links to many of the groups Phelps recommended we cite, including Achieve, the Fordham Foundation, The Education Commission of the States, the National Goals Panel."

The only one of these four I listed is Achieve. I mentioned 22 other organizations which EWA does not list, and that's not counting the many test development firms, or their common PR group. Moreover, my list was just one I thought up off the top of my head. I include here the names of some more organizations which either favor testing or feel neutral about it, but still know something about it. I list them at the bottom.

Phelps ignored other work EWA has done on testing

True, all I did was look at their website and, even then, just the summer 2000 version of the website. What I did, however, was similar to what any reporter would have done this summer who was not among the 250 who attended the EWA national meeting or those who attended EWA's other 3-day meeting to which Ms. Walker refers (i.e., the vast majority of reporters in this country). Indeed, I probably spent more time at the website than your average reporter would have who assumed that anything she was reading was thorough and balanced and could be trusted.

"The irony, of course, is that the opponents of testing that he accuses EWA of being biased toward complained about just the opposite - that we represent the proponents of testing too often in our work."

Sure, some testing opponents would rather EWA didn't mention the proponents or the neutrals at all, which it has come very close to doing at its website. Some of the anti-testing organizations have a lot of money and other resources. They have the time to troll the web and check on every education site to see if it is presenting their propaganda and, likewise, not presenting any opposing views. And, there are many websites that purport to represent all research on testing but provide links exclusively to anti-testing groups.

I think it would be a terrible policy for EWA to use testing opponents' complaints, or my complaints for that matter, as benchmarks to judge the balance of its coverage, however. I'm reminded of the story of the opportunist who sues someone he doesn't like on a phony charge in a court with a lazy judge. The judge doesn't want to make the effort to weigh the evidence and so just splits the difference. The opportunist gets half of what he sued for. So, he sues again on a phony charge and asks for even more. Again, he gets half of what he asked for. I forget how the story ends, but just this much makes the point.

"The majority of what journalists quote in a newspaper is pro-testing." (Diane Loupe).

Whereas, for all I know, that may have been true in the past, I believe it has not been true for at least a few months. This summer, I have made daily downloads of all articles on testing I could find at EducationNews.org, Education Week, and any other sources, read them, and counted them up. (My wife resents the huge pile of paper that has accumulated in our den.) The large majority of articles on testing don't take any side; they're just factual. The balance among the rest, for this summer at least, and judging mostly based on who the reporters chose to talk to, has been strongly anti-testing.

"Mr. Phelps is correct that most education journalists don't read scholarly journals, but I'll bet most education writers spend more time in classrooms than a lot of education scholars and may be more in touch with the reality of schools in their area." (Diane Loupe)

It wouldn't surprise me if Ms. Loupe is correct that reporters spend more time in classrooms than many education professors and non-professorial education scholars. I raised the issue of journals because that is where the most knowledgeable testing experts conduct their discussions. One won't find them in advocacy groups. Though, granted, not all academic journals are unbiased or balanced, particularly in education. The more technical journals, however, usually are pretty trustworthy; at least I think that's true in testing. And those journal editors, or members of their editorial board, would be good sources to consult on testing issues. Indeed, those are exactly the kind of people that EWA should be putting reporters in contact with, not the usual crew of ax-grinders.

This is a pertinent issue because, at least it seems to me, journalists who discuss the *research* on testing tend to rely on anti-testing advocacy groups' representation of it, and that representation tends to be not particularly balanced and, often, is just plain wrong. Rarely have I read a journalist questioning the validity of the research presented in this way and attempting to get an opposing point of view.

Finally, I cannot read minds and I do not know the motives of the EWA staff. Maybe it is

not *deliberately* biased, maybe it is or maybe it is not naive. I can't say. But, in my judgement, its *website* has been biased *this summer* on the topic of *testing*.

All should realize that the debate on testing (high-stakes testing, that is) is not limited to a world of tweed jackets, sweet-smelling pipes, and green quadrangles where nice, earnest professors have polite disagreements with each other. It's part of a war for the control of our country's schools, being fought by the insiders against outsiders. The booty is our children's futures. The stakes are enormous. The question is: is the EWA up to covering a war, or not? If it is, it should know what the fight is about, who the combatants are on both sides, and it should spend some time on both sides.

Richard P. Phelps is the author of *Education Establishment Bias? The National Research Council's Critique of Test Utility Studies*, *Why Testing Experts Hate Testing*, and *Test Basher Benefit-Cost Analysis*

http://www.siop.org/tip/backissues/Tipapr99/4Phelps.htm

http://www.edexcellence.net/library/phelps.htm and

http://www.edexcellence.net/library/issuespl/subject/standar/testbash.html

P.S. Between the time I first looked at the EWA website and now (Sept.1), EWA added a link and a blurb for statistician David Murray's 1999 article *The War on Testing* in the "Reporters' Stories" section. That's a start.

Other organizations knowledgeable about standardized testing who could serve as resources for reporters:

Southern Regional Education Board (SREB)

Center for Advanced Human Resource Studies (CAHRS), Cornell University

National Association of College Admission Counselors (NACAC)

Office of Public and Governmental Affairs, CTB Macmillan-McGraw-Hill

National Education Consumers' Clearinghouse

CEO America

United States Chamber of Commerce

The Business Roundtable

American Council on Education

Law School Admissions Council

Board of Bar Examiners (any state)

Medical College Admissions Council

Board of Medical Examiners (any state)

American College Testing

National Organization for Competency Assurance

Personnel Decisions Research Institute

National Center on Education and the Economy

Society for the Advancement of Excellence in Education (Canada)

http://www.dmoz.org/Reference/Education/Educational_Testing

Post your comments on this article

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

Shop Here for Books
amazon.com

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 8

Walt Haney's Texas Obsession, Part 1: SAT Scores

Richard P. Phelps

Among the small group of testing opponent/education professors getting so much attention from the press down in Texas these days is one from Boston College, Walter Haney. He has described the "Texas miracle" as a "mirage," and the Texas Assessment of Academic Skills (TAAS) as a "sham." His article in the *Education Policy Analysis Archives* (EPAA) can be found further down the list of commentaries and reports, or at:

http://olam.ed.asu.edu/epaa/v8n41/part7.htm

I confess that I am intrigued by Walt Haney's research, not because of his studies' conclusions, which I believe are pretty predictable, but by the manner in which he gets to his conclusions. I may be one of the few who reads his work in detail, however. He throws many numbers at the reader and often meanders and digresses as he, reasonably, tries to explain the relevance and character of each new clump of data that he adds to the mix. One cannot clearly see how he builds his superstructure of numbers and facts into a complete story without committing some time to it; probably most readers just cross their fingers and assume (or hope) that the conclusions derive from his evidence. Those readers are missing out, I think, because tracking down his numbers and trying to verify his conclusions can be a lot of fun.

I intend to devote a few of these commentaries to Walt Haney's critique of the TAAS, which press accounts say he spent two years studying. This week, I focus on his analysis of Scholastic Assessment Test (SAT) scores and his comparison of SAT score trends in Texas to TAAS score trends (section 7.3 of his EPAA article). In later commentaries, I will look at his similar analysis of National Assessment of Educational Progress (NAEP) scores and his estimates of Texas dropout numbers.

Texas's rank in SAT state average scores

All Texas students in certain grade levels are required to take the TAAS, but only students who wish to enter college need take the SAT. Only about 50% of Texas' high school students end up taking the SAT, most in their senior year. The College Board, the organization responsible for the SAT, released this year's results, with averages for the nation and states, just two weeks ago.

Some critics remarked on how low Texas's average SAT scores were relative to other states' scores (40th in math, 46th in verbal), pretty low for an "education miracle" state.

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW
Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests

As Donna Gardner wrote in this space 13 days ago, "Please correct me if I am wrong. According to the just-released SAT scores, Texas' verbal score was the third worst in the U. S., behind South and North Carolina. If memory serves me correctly, North Carolina and Texas have been touted as two of the nation's most outstanding leaders in education reform. Something does not add up."

The quickest rejoinder to this criticism is to say that Texas's low standing on the SAT is hardly a product of the TAAS era, but has existed for a long time. As one can see in Haney's graphs of SAT average scores for Texas and the U.S. from 1970 to 1999 (see below), the gulf between the Texas and the U.S. SAT mean scores has been about 10 scale points for verbal and 12 for math. These TAAS-era differences are smaller than those of the mid-1980s, long before TAAS, of 13 for verbal and 16 for math. So, it would appear that TAAS is not to blame for Texas's low SAT average.

Here's a somewhat longer rejoinder. As far as the SAT is concerned, there are three groups of test-takers who influence the level of each state's average score. **First**, there are those students in a state who do not plan to go to college and who do not attempt the SAT. They tend to be lower achieving students who would probably bring the state average down if they did take the SAT.

Second are those students who take the SAT because their relatively low-cost and/or nearby state public higher education institutions require the SAT. This scenario plays out in less than half the U.S. states, however. That's because a slight majority of the states (or, rather, their public higher education institutions) prefer or require the SAT's competitor, the American College Test (ACT). This second population of SAT test-takers does exist in Texas because its public higher education institutions require the SAT. These test-taking students tend to score higher on the SAT than the first group of non-college-bound probably would, as these college-bound are higher-achieving students.

Again, this second population of test-takers does not exist in the majority of U.S. states. A wide swath of territory from the Sierra Nevada to the Appalachians (with the exceptions of Texas and Indiana) encompasses ACT states. Students who take the SAT in these states are students who plan to attend college outside their own state, where they will pay out-of-state tuition, private school tuition, or require a scholarship or grant. These students must be very high-achieving students to get into these institutions in other states which

Ideas Practicas Para Padres de Familia pdf file

Special Education Resources

TAD for English/Language Arts/Reading

LA Times Special Report Special Education a Failure on Many Fronts

AVKO

Parental Opt-Out Form

Public Policy Org

NEW Publications francaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs

have no obligation to take them. This **third** population of test-takers tend to produce the highest SAT average of all. Thus, one finds the highest average SAT scores in ACT states, such as North Dakota, where the state average SAT math score is 609, more than one standard deviation above the national mean.

It simply is not fair to compare the average SAT score in an SAT state to one in an ACT state, as was done implicitly in Education Secretary William Bennet's famous (or infamous) "wall chart" in the early years of the Reagan administration.

If one compares Texas's SAT average scores only to those of other SAT states, Texas's rank doesn't look so awful. Twelve states have higher math and verbal scores. Seven states have one score (math or verbal) higher than Texas's, but not the other. Four states have lower math and verbal scores. Given Texas's relatively large population of limited-English-proficient high school students (second only to California's), this ranking does not seem so dreadful. Test critics are often quick to pull out SES differentials as an excuse when they don't like test-score comparisons, so they should not begrudge this one.

Haney makes a much narrower comparison than I do, of Texas to only seven other states with SAT student participation rates between 49 and 53 percent. In his comparison, Texas looks worse. The interval of 49 to 53 percent seems rather contrived to me.

The contrast of Texas's flat SAT scores in the 1990s with the TAAS's rising scores

While Haney chides Texas for its low relative standing on SAT means, he also observes that SAT scores for Texas are not rising in conjunction with TAAS scores, which makes him suspicious of the validity of the TAAS score rise.

He writes: "...it is relevant to address the question of whether gains on TAAS are a real indication of increased academic learning among students in Texas or whether they represent scores inflated due to extensive preparation for this particular test.

"To help answer this question, it is necessary to look at other evidence of student learning in Texas, to see whether the apparent gains on TAAS since its introduction in 1991 are reflected in any other indicators of student learning in Texas.

"Suffice it to say that the general conclusion of these analyses is that, at least as measured by performance on the SAT, the academic learning of secondary school students in Texas has not improved since the early 1990s, at least as compared with SAT-takers nationally....the pattern of results on both the SAT-V and SAT-M for Texas secondary school students relative to students nationally fails to confirm the gains on the exit level TAAS during the 1990s."

In response to Haney's compelling argument, I state two facts: 1) The SAT and the TAAS are different tests. The SAT, originally designed to measure "aptitude," is as wide-ranging and general as an academic test can be, whereas the TAAS is based on specific curriculum standards agreed to by the citizens of Texas. Moreover, the SAT level of difficulty is defined by the achievement level of the average college aspirant, whereas the TAAS is, by many accounts, pitched to a minimum level of achievement of something like a 6^{th}- or 7^{th}-grade level of difficulty.

Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

(2) More important to the point, the population of students taking each test is different, too - college aspirants for the SAT and the entire range of students for the TAAS.

A decade ago, a PhD economics student at MIT, Jonathan E. Jacobson, completed a dissertation on the effect of 1980s-era minimum competency tests on academic achievement and young adult earnings. He used the National Education Longitudinal Study (NELS), which consists of a panel group questionnaire with an achievement test imbedded. He compared the most recent average NELS achievement test scores and subsequent respondent earnings between those individuals who had attended high school in minimum competency test states and those who had not.

As expected, he found that the lowest-achieving students were, indeed, made better off in the minimum-competency test states. Both their average test scores and their average level of earnings were substantially higher than those of their counterparts, the lowest-achieving students from non-test states. This was encouraging. However, he also found that the other students, particularly the middle-achieving ones, in the minimum-competency testing states ended up worse off. Their average NELS achievement test scores and their subsequent average earnings were lower than those for their counterparts, the middle-achieving students from non-test states.

It is easy to speculate about what might have happened to produce these results. In test states where the only threshold to exceed was that at the minimum level of achievement, substantial effort focused on that level. Perhaps schools and teachers kept themselves so busy making sure the lowest-achieving students got over the minimum competency threshold that they neglected the higher-achieving students and the curriculum and instruction that they needed.

Given this evidence, and given the low level of the TAAS, one would expect that the average SAT scores of college-aspiring Texas high schoolers would be *declining* in the TAAS era, instead of remaining flat. There is no logical reason why an instructional focus on a 6^{th}- or 7^{th}-grade level TAAS would help college aspirants on their SATs. The only effect that one would expect would be a negative one, with instructional resources flowing away from the higher-achieving students toward the lower-achieving students.

The fact that Texas's SAT scores have remained flat throughout the TAAS era (that is, the same distance in scale points below the national average), is a credit to the Texas testing program. It apparently has not degraded instruction for its higher-achievers, as minimum competency testing regimes can do.

One solution to a "Jonathan Jacobson effect," by the way, is to provide something for the higher-achieving students to aim for, too, that will inspire them to higher achievement. Many other countries offer more than one high-stakes test at different levels of difficulty in their testing programs. Texas, for its part, could offer an "honors diploma," in addition to its regular diploma, that would require passage of a test more difficult than the TAAS.

The curves that average SAT scores trace over time

Haney draws line plots of U.S. and Texas average SAT scores for the time period 1972 to 1999 (see below). Why he went all the way back to 1972, I don't know, but the wide range

of years is illustrative. One can observe that the precipitous decline in the relative standing of Texas on the SAT occurred between the mid-1970s and the mid-1980s. Since then, and including during the TAAS era, the distance between the Texas and U.S. averages has been relatively constant. It might be interesting to investigate what might have happened from about 1975 to 1985 to induce the SAT score decline, but, most assuredly, it was not the TAAS.

One can see in the SAT-Verbal graph that the mean-score differential between Texas and the United States has remained virtually constant throughout the TAAS era. For each year from 1990 through 1999 the difference was either 10 or 11 scale points.

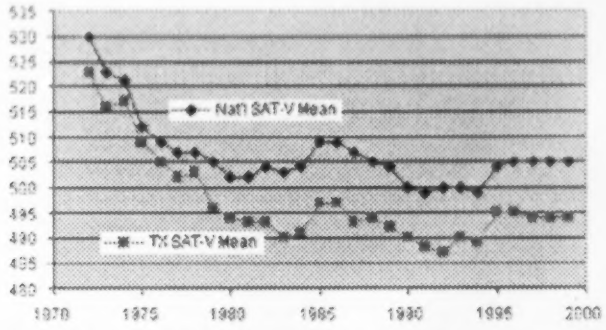


Figure 7.4 Average SAT-V Scores, Texas and National, 1972-1999

The SAT-Math graph displays a somewhat different relationship in the 1990s. In 1990, the scale score difference between Texas's average and the U.S.'s average was 12 points. In 1999, the difference was, again, 12 points. No change. Haney tries to make a large point about the curvature of the Texas line plot during the in-between years, however: "...the pattern of results on the SAT-M indicates that at least since 1993, Texas students' performance on the SAT has worsened relative to students nationally [from a 5- to a 12-point difference]."

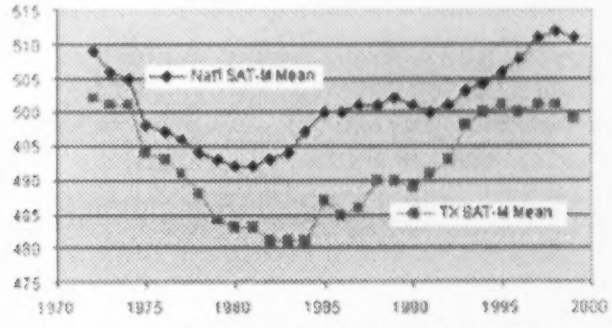


Figure 7.5 SAT-M Scores Texas and National, 1972-1999

To be complete, he would need to add that Texas students' performance on the SAT improved relative to students nationally in the years 1990 to 1993, also TAAS years, from a 12- to a 5-point difference.

BEST COPY AVAILABLE

Is there enough variation in that little curve to set up a statistical test? Remember, these graphs are zoom-lens shots of much larger graphs. Seven scale points here may seem a lot; on a graph with a range from 200 to 800 points (the SAT range), it might look trivial.

Maybe I'm wrong, though, and there is a story there. What might it be? Could the Texas math curve be evidence of a Jonathan Jacobson effect, perhaps? Maybe so. Maybe Walt Haney could do that study, too.

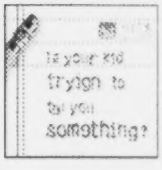
Richard P. Phelps is the author of *Why Testing Experts Hate Testing* and *Test Basher Benefit-Cost Analysis*.

http://www.edexcellence.net/library/phelps.htm

http://www.edexcellence.net/library/issuespl/subject/standar/testbash.html

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF



CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us
Advertising
Archives
DailyNews
Past Five Days
Book Reviews
College Aid
Book Reviews
College Aid

Test Bashing, Part 9

Walt Haney's Texas Obsession, Part 2: NAEP Scores

Richard P Phelps

Walt Haney has argued that trends in Texas Assessment of Academic Skills (TAAS) scores should be reflected in Texas' SAT scores. Last week, I disagreed and argued that, if anything, SAT trends reinforced a conclusion that the TAAS was a good program. Haney further argues that correlations between TAAS score trends and National Assessment of Educational Progress (NAEP) trends are bogus because Texas manipulated NAEP participation levels, excusing an artificially high number of limited-English-proficient students. This week, I write that Haney used the wrong numbers, otherwise miscalculated NAEP's and others' statistics, and might have just made some things up. NAEP score trends do, indeed, affirm the TAAS score gains.

As most readers in this venue probably know, the NAEP is a "no-stakes" test funded by the federal government that, every two years, tests 4th, 8th, and/or 12th graders in national samples of schools on their knowledge of one or more of several major academic subjects. On some occasions, the NAEP also tests samples within states that wish to participate. Most do, NAEP state assessments usually involve 40 or more states.

NAEP scores trends can be used as an unbiased check on TAAS score trends. If the TAAS score rise was for real, they should be reflected to some degree in NAEP score trends, though, one should not expect a perfect correlation as there are many differences between the two tests. To a large degree it's a valid assertion, however, because the TAAS and NAEP are both curriculum-based tests of major subject areas, such as math and English. If the TAAS score increase over the 1990s were not paralleled in state NAEP score trends for Texas to some degree, one might reasonably wonder if there might be something misleading about the TAAS score increases.

In Haney's words (which I highlight in green): "In 1997, results from the 1996 the National Assessment of Educational Progress (NAEP) in mathematics were released The 1996 NAEP results showed that among the states participating in the state-level portion of the math assessment, Texas showed the greatest gains in percentages of fourth graders scoring at the proficient or advanced levels. Between 1992 and 1996, the percentage of Texas fourth grades scoring at these levels had increased from 15% to 25%. The same NAEP results also showed North Carolina to have posted unusually large gains at the grade 8 level, with the percentages of eighth graders in North Carolina scoring at the proficient or advanced levels improving from 9% in 1990 to 20% in 1996. (Reese et al., 1997)

" ... these findings led to considerable publicity for the apparent success of education

BEST COPY AVAILABLE

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

reform in these two states. The apparent gains in math, for example, led the National Education Goals Panel in 1997 to identify Texas and North Carolina as having made unusual progress in achieving the National Education Goals."

http://www.negp.gov/reports/grissmer.pdf

David Grissmer, of the Rand Corporation, and the author of that study for the National Goals Panel later expanded his scope to include all the U.S. states, and came to the conclusion that states with high-stakes testing programs seemed to be improving student achievement, as measured by the state NAEP, at a faster pace than other states, everything else held equal.

http://www.rand.org/cgi-bin/Abstracts/ordi/getabbydoc.pl?doc=MR-924

As Haney writes: "...the gains on TAAS...appear quite impressive. Across all three grades and all three TAAS subject areas (reading, math and writing), the magnitude of TAAS increases ranged from 0.43 to 0.72 standard deviation units. According to guidelines for interpreting effect sizes, these gains clearly fall into the range of medium to large effects. Also, the gains on TAAS clearly exceed the gains that appear possible, according to previous research, from mere test coaching. The gains on TAAS seem especially impressive when it is recalled that the gains on TAAS. . represent performance of hundreds of thousand of Texas students..."
"Apparent gains for Texas in NAEP math scores between 1992 and 1996 were indeed statistically significant.... Also ..the NAEP math gains for Texas fourth graders between 1992 and 1996 were greater than the corresponding gains for any other state participating in these two NAEP state assessments. So any reasonable person must concede that the apparent improvement of Texas grade 4 NAEP math average from 217.9 in 1992 to 228.7 in 1996 (a gain of about one-third of a standard deviation), if real, is indeed a noteworthy and educationally significant accomplishment."

Haney then goes on to describe the evidence that proves, in his opinion, that "the apparent improvement" in academic achievement in Texas is not "real." His evidence consists of the following:

1. retention rates prior to grade 4 (Haney argues that Texas holds more students back in the primary grades and, thus, its students taking the grade 4 NAEP are older than other states'.);
2. rates of participation by students with disabilities (SD) and limited English proficiency (LEP) in the state NAEP (Haney argues that Texas excused more SD and LEP students from taking the NAEP, thus raising the NAEP scores for Texas, if one assumes that SD and LEP students would score low.);
3. level of effort to comply with NAEP's new criteria, as of 1996, for inclusion of SD and LEP students in the state NAEP (Haney argues that Texas did not cooperate with the national effort to include more SD and LEP students in the state NAEP and, instead, excused even more of them in 1996 than it had in 1992, thus biasing the trend of its NAEP scores upward.); and
4. relative average of state NAEP scores (Haney repeatedly points out that, regardless of what one thinks about trends in state NAEP scores, the average NAEP scores for Texas (and North Carolina) are not exemplary, they rest only near the national average.

BEST COPY AVAILABLE

Parental Opt-Out Form

Public Policy Org

NEW Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research
The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

The relevant passages of Haney's article can be found in:

http://olam.ed.asu.edu/epaa/v8n41/part7.htm

1. Retention rates prior to grade 4

Haney writes: ".. NAEP state assessments have focused on measuring the learning of students at particular grade levels, namely grades 4 and 8. This constitutes a little recognized limitation of NAEP, viz., that in focusing on performance of students enrolled in grades 4 and 8, results of NAEP state assessments are inevitably confounded with grade retention differences across the states. This means that in states in which failure and grade repetition are common, students in grades 4 and 8 will be older than students in states where grade retention is less common. Thus, it is probably no accident that the two states identified in 1997 by the NEGP as having made unusual "progress" on NAEP math assessments, Texas and North Carolina, have unusually high rates of failure and grade repetition before grade 4 (see Heubert & Hauser, Table 6-1, corrected)."

Haney is implying that Texas and North Carolina 4th graders do as well as they do, on average, only because so many of them are given extra time prior to 4th grade to learn. They are given the extra time - more time than students in other states get - because Texas and North Carolina hold more students back in the primary grades than other states do. To check Haney's claims, I looked at the Table 6-1 in the Heubert and Hauser book (pp. 138-147 in the corrected form of an errata sheet), and added up the retention rates for the primary grades (grades 1 through 3) for all the states listed in the table. I do not get Haney's results.

Out of the 19 states listed , North Carolina ranks 7th, with an 11.3 percent cumulative retention rate for grades 1-3. Texas ranks 16th out of 19, with a 5.1 percent cumulative retention rate.

If I add in kindergarten, I lose three states without data for kindergarten. After calculating the cumulative percentage retention rate for each state for grades K-3, I find much the same result as before. North Carolina ranks 5th out of 16 states, with a 15.5 percent cumulative retention rate, whereas Texas ranks 10th out of 16 states, with an 11 percent cumulative retention rate.

Texas and North Carolina do not have "unusually high rates of failure and grade repetition before grade 4," according to the table Haney refers us to. Indeed, Texas's rate is well below average. I encourage the reader to look at the table and check my addition. The report is *High Stakes*, by the National Research Council's Board on Testing and Assessment, Jay P. Heubert and Robert M. Hauser, Editors. It is available on line at http://www.nap.edu

Incidentally, a review of Heubert and Hauser's rather one-sided report, *High Stakes*, will appear in the December 2000 issue of *Educational and Psychological Measurement*.

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

2. Exaggerating the increase in NAEP scores between 1992 and 1996 - "Illusion from Exclusion"

Haney writes: "Apparent gains for Texas in NAEP math scores between 1992 and 1996 were indeed statistically significant. Also NAEP math gains for Texas fourth graders between 1992 and 1996 were greater than the corresponding gains for any other state participating in these two NAEP state assessments. So any reasonable person must concede that the apparent improvement of Texas . , if real, is indeed a noteworthy and educationally significant accomplishment.

"But there is that 'if' The other perspective not yet brought to bear in considering changes in NAEP test score averages is advice offered in Part 1. When considering average test scores, it is always helpful to pay attention to who is and is not tested.

Table 1
Percentages of IEP and LEP Students
Excluded from NAEP State Math Assessments, Texas and Nation

Mathematics, Grade 4	**1990**	**1992**	**1996**
Texas		8%	11%
Nation		8%	6%
Mathematics, Grade 8			
Texas	7%	7%	8%
Nation	6%	7%	5%

Note: "IEP" = "individual education plan" (i e. student with disability)

Source: Reese et al., 1997, pp. 91, 93; Mullis et al., 1993, pp. 324-25

"As can be seen in this table [reproduced above as table 1], at the national level, between 1992 and 1996, the percentages of students excluded fell slightly (from 8% to 6% at grade 4, and from 7% to 5% at grade 8). These results at the national level were presumably a result of efforts to make NAEP more inclusive in testing LEP and special education students. However, in Texas, the percentages of students excluded from testing increased at both grade levels: from 8% to 11% at grade 4, and from 7% to 8% at grade 8. This means that some portion of the increased NAEP math averages for Texas in 1996 are illusory, resulting from the increased rates of exclusion of LEP and special students in Texas from NAEP testing. The gaps in rates of exclusion between Texas and the nation in 1996 also mean that comparisons of Texas with national averages in that year will be skewed in favor of Texas for the simple reason that more students in Texas were excluded from testing. In short, as with TAAS results, some portion of the apparent gains on NAEP math tests in Texas in the 1990s is an illusion arising from exclusion."

BEST COPY AVAILABLE

Haney's argument is compelling, but his numbers are wrong. He either doesn't realize, or he isn't telling us, that the national numbers he cites and the Texas numbers he cites come from completely different samples with completely different trends. The national numbers are, apparently, from the NAEP national sample used for national trend data. It contains items that match items used in the past so that NAEP can calculate valid trends in student achievement over time. The state numbers come from state NAEP samples, and the average of the state samples is not necessarily the average of the national sample Haney uses. (Don't ask me to explain why the percent excluded in the national trend sample is different than the average percent of all the state samples. You'd have to ask NAEP.)

The reader is welcome to look at the table and check. I found Table B.4 in the *1992 NAEP Trial State Assessment Data Compendium*, pages 796-797 for the 1992 exclusion percentages and in the *NAEP 1996 Mathematics Report Card for the Nation and the States*, Table D.2, Page 144. for the 1996 percentages. First I get the percentages for Texas (1992: 8% (grade 4) and 7% (grade 8); 1996: 10% (grade 4) and 9% (grade 8)). I have no idea where Walt Haney got his 11 and 8% for Texas for 1996.

The 1996 NAEP Almanac can be found on the web:

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=97488

Next, instead of using the exclusion percentage listed for the "Nation," which is from the wrong sample, I calculate the (unweighted) average of all the state percentages. Try it yourself (note that you need to use the "S1" column from the 1996 table because the NAEP scores were calculated using those inclusion criteria). I calculate the following as state averages: (1992: 5.1% (grade 4) and 5.2% (grade 8); 1996: 7.2% (grade 4) and 6.3% (grade 8)). I put all these numbers in Table 1b below.

Table 1b
Percentages of IEP and LEP Students
Excluded from NAEP State Math Assessments, Texas and State Averages

Mathematics, Grade 4	**1992**	**1996**
Texas	8%	10%
States' Average (unweighted)	5.1%	7.2%
Mathematics, Grade 8		
Texas	7%	9%
States' Average (unweighted)	5.2%	6.3%

Source: Reese et al., 1997, pp. 91, 93; Mullis et al., 1993, pp. 324-25

Haney's disparity disappears. The trend between 1992 and 1996 for grade 4 is in the same direction for Texas as it is for the states as a whole. The trend for grade 8 is also in the same direction for Texas as it is for the states as a whole. Indeed, all exclusion percentages are trending up between 1992 and 1996. Moreover, the magnitude of the trends for Texas and for all the states are much the same, within rounding error.

The percentages in Table 1b make much more sense than those in Haney's table. Notice that in Haney's table (table 1 above), Texas and the Nation have exactly the same exclusion percentages in 1992 for grades four and eight. Anyone familiar with the situation would know that that cannot be. Texas has the second highest proportion of limited English proficient students in the country, second only to California, and its proportion of students with disabilities is not extraordinarily low. Logically, Texas must have a higher exclusion percentage than the country as a whole. Indeed, given its geographic position, it should only be surprising that Texas's exclusion percentage is not higher than it is.

Incidentally, using Haney's method of comparison, one could make any state, not just Texas, look like it was slack in its effort to decrease exclusion rates.

It might also be worth noting, as it relates to motive, that Haney does not mention the continuing high rate of immigration into Texas between 1992 and 1996 in his discussion of the exclusion issue (more LEP students would justify an increase in the exclusion rate). He mentions it in other locations in his article, where it serves his purpose. Here, in this discussion, it would counter his argument and, perhaps, that's why he left it out.

3. Excluding SD and LEP students from NAEP

Haney writes: "Because excluding sampled students from NAEP testing has the potential for skewing results, over time NAEP has developed detailed guidelines for excluding students from testing, and has taken special steps to try to include LEP and special education students in NAEP testing, for example, by allowing accommodations to standard NAEP testing procedures to meet the needs of special education students.

"As can be seen in this table [table 1 above], at the national level, between 1992 and 1996, the percentages of students excluded fell slightly. ... These results at the national level were presumably a result of efforts to make NAEP more inclusive in testing LEP and special education students. However, in Texas, the percentages of students excluded from testing increased at both grade levels. .. This means that some portion of the increased NAEP math averages for Texas in 1996 are illusory, resulting from the increased rates of exclusion of LEP and special students in Texas from NAEP testing....the apparent gains on NAEP math tests in Texas in the 1990s is an illusion arising from exclusion."

Haney is correct in saying that NAEP encouraged states to include more SD and LEP students in their NAEP testing. Moreover, NAEP also introduced new criteria to guide the states toward doing so.

Haney either doesn't realize, or he isn't telling us, however, that the published NAEP scores for 1996 are based on the old inclusion criteria. So, his entire digression about the new criteria and his allegation that Texas was flaunting the new criteria is meaningless for explaining score trends.

Since he has leveled the rather slanderous charge, however, that Texas deliberately excludes too many SD and LEP students from NAEP testing and that it neglects to adhere to NAEP exclusion criteria, let us look into it.

BEST COPY AVAILABLE

There is a very revealing table in the *1996 NAEP Almanac* that Haney either doesn't know about, or isn't revealing to us. Table D.4 lists the percentage of limited English proficient (LEP) students included in 1996 NAEP testing by each state with a population of LEP students large enough to provide an adequate sample for estimating summary statistics.

According to the original NAEP inclusion, Texas included (i.e. tested) 66 percent of its 4th-grade LEP students, which places it 4th out of 10 states with large LEP populations, and 55 percent of its 8th- grade LEP students, which places it 1st out of four states, ahead of Arizona, California, and New Mexico. Texas's relative position looks even better when calculated according to the newer inclusion criteria.

Not only is the state of Texas not responsible for all the devious and unethical behavior of which Haney accuses them, they are, in fact, one of the best and most reliable states in regard to including their LEP students in NAEP testing.

The dedicated, hard-working employees of the Texas Education Agency deserve better than the treatment they're getting from Haney. I can understand that the vested interests in education are attacking Texas because they feel threatened by the policies of its governor, who is running for president. But, in the manner in which they are doing it, they also happen to be trampling on the reputations of thousands of sincere and competent education professionals in Texas's education department. That's not right.

4. Texas's NAEP scores are just average

As if what Haney did that is described above isn't bad enough, he also chides Texas (and North Carolina) for being just average in their NAEP scores: "...review of results of NAEP from the 1990s suggests that grade 4 and grade 8 students in Texas performed much like students nationally. On some NAEP assessments, Texas students scored above the national average, and on some below. In the two subject areas in which state NAEP assessments were conducted more than once during the 1990s, there is evidence of modest progress by students in Texas; but it is much like the progress evident for students nationally."

Haney joins Linda McNeil in making this grumpy criticism. Texas (and North Carolina) used to be at the bottom of the pack. Now, they have risen to the middle. Instead of rejoicing at their progress, Haney and McNeil can only muster up resentment.

Conclusion

Giving Walter Haney the last word: "...the magnitudes of the gains apparent on NAEP for Texas fail to confirm the dramatic gains apparent on TAAS. Gains on NAEP in Texas are consistently much less than half the size (in standard deviation units) of Texas gains on state NAEP assessments. These results indicates that the dramatic gains on TAAS during the 1990s are more illusory than real. The Texas "miracle" is more myth than real."

BEST COPY AVAILABLE

But, as Haney himself writes, "At this point, the reader may begin to doubt the consistency of my approach to data analysis."

Richard P. Phelps is the author of *Test Basher Arithmetic*, *Test Basher Benefit-Cost Analysis*, and *Why Testing Experts Hate Testing*:

http://www.edweek.org/ew/1998/26phelps.h17

http://www.edexcellence.net/library/issuespl/subject/standar/testbash.html

http://www.edexcellence.net/library/phelps.htm

Post your comments on this article

Test Bashing, Part 9
Walt Haney's Texas Obsession, Part 2: NAEP Scores
Richard P. Phelps

Test Bashing, Part 8
Walt Haney's Texas Obsession, Part 1: SAT Scores
Richard P. Phelps

Test Bashing, Part 7
The Unfortunate Bias of the Education Writers' Association (continued)
Richard P. Phelps

Test Bashing, Part 6
The Education Press's Cop-Out on Testing (continued)
Richard P. Phelps

The Computer's Burden
Education Polls
Richard Phelps

Test Bashing, Part 5
More on Texas Testing and School Reform
Richard P. Phelps

Test Bashing, Part 4
That "Backlash" That Testing Opponents So Desperately Crave
Richard P. Phelps

Test Bashing, Part 3
The Education Press's Cop-Out on Student Testing
by Richard P. Phelps

Test Bashing Texas, Part 2:
The Research Sez...Standardized Tests are Horrible and Terrible!!!

Test Bashing Texas, Part 1:
The *Washington Post* Advocates State Student Testing Programs, But Not in Texas

by Richard P. Phelps

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUE

CALL FOR PAPERS
In Defense of Testing
Test Bashing Series
Intellectual Property and Digital Information in Higher Education
LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report
Shop Here for Books
amazon.com
CAPAlert for Parents & Grandparents
About Us
Advertising
Archives
DailyNews
Past Five Days
Book Reviews
College Aid
Book Reviews
College Aid

Test Bashing, Part 10

Walt Haney's Texas Obsession, Part 3: Texas as Pariah State

Richard P. Phelps

Walt Haney argues that correlations between TAAS and NAEP score trends are bogus because Texas manipulated NAEP participation levels, excusing an artificially high number of LEP students. Last week, I showed that Haney used the wrong numbers, otherwise miscalculated statistics, and apparently just made some things up. NAEP score trends affirm the TAAS score gains. This week I turn to Haney=s indefatigable insistence that education in Texas is worse than practically anywhere, and that any evidence to the contrary must be an artifact of gross dishonesty on the part of Texas education professionals.

Haney makes many comparisons of Texas to the United States as a whole, all showing Texas to be worse. He also prints several state rankings in his report, each showing Texas to be near the bottom, and getting worse, according to one or another criterion.

What he does not do is compare Texas to its peers. Nor does he attempt to adjust for Texas=s unique demographic characteristics in making his comparisons. The latter behavior is particularly ironic because much of Haney=s rationale for his anti-TAAS crusade is an alleged interest in the well-being of Texas=s minority students. Throughout his report, he breaks out statistics and test scores by ethnicity in order to demonstrate the variant impact on the different groups.

By Haney=s logic, Texas has no cause for celebration if its students= achievement improves unless the test score gap between whites and blacks narrows...not even if blacks= scores improve significantly. His demographic imperative is so dominant that no success can be claimed when minority students do well, unless white students also do less well.

Yet, when it serves his purpose, he insists that for the state of Texas to be genuinely successful, it must match or exceed the norm for the nation as a whole. No matter that it has a long history with a vast poor yeoman class, a large poor minority population, and rapid, continuous immigration, both legal and illegal, of mostly unskilled workers from south of its border. Haney insists that Texas bear all the marks of a Connecticut or a Wisconsin or he will drub it to shame with insults and slander.

Below, I examine four cases where Haney disparages Texas in comparing it to the nation or the other states. First, I briefly summarize the NAEP score trends discussion of last week and, then, I look at three different measures that Haney claims demonstrate that Texas has an extraordinarily large dropout problem. I introduce each case first with

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

Haney=s statistical claim (with Haney=s writing in green) and then I compare Texas, using Haney=s numbers, not to the nation as a whole or to all the other states, but to a collection of states with similar geographic and demographic characteristics.

1. Excluding Limited-English-Proficient Students from NAEP Testing

A...at the national level, between 1992 and 1996, the percentages of students excluded [from NAEP testing] fell slightly.... These results at the national level were presumably a result of efforts to make NAEP more inclusive in testing LEP and special education students. However, in Texas, the percentages of students excluded from testing increased at both grade levels.... This means that some portion of the increased NAEP math averages for Texas in 1996 are illusory, resulting from the increased rates of exclusion of LEP and special students in Texas from NAEP testing ...the apparent gains on NAEP math tests in Texas in the 1990s is an illusion arising from exclusion.@

There is a very revealing table in the *1996 NAEP Almanac* that Haney either doesn=t know about, or isn=t revealing to us. Table D.4 lists the percentage of limited English proficient (LEP) students included in 1996 NAEP testing by each state with a population of LEP students large

enough to provide an adequate sample for estimating summary statistics.

According to the table, Texas included (i.e. tested) 66 percent of its 4th-grade LEP students, which places it 4th out of 10 states with large LEP populations, and 55 percent of its 8th- grade LEP students, which places it 1st out of four states, ahead of Arizona, California, and New Mexico.

Not only is the Texas Education Agency not responsible for all the devious and unethical behavior of which Haney accuses it [of deliberately excluding LEP and SD students in order to artificially boost Texas=s average test scores], it is, in fact, one of the best and most reliable state education agencies in regard to including LEP students in NAEP testing.

2. Dropouts and High School Completion Rates

AEven if we use the very conservative estimates of high school completion derived from CPS data (and reproduced in Table 7.2 below) we see that Texas has a rate of non-completion of high school among young adults of about 20%CCmore than 5 percentage points above the national rate.

ATable 7.2 reproduces a table from the latest NCES dropout report, showing high school completion rates of 18 through 24 year-olds, not currently enrolled in high school or below, by state: October 1990-92, 1993-95 and 1996-98. As can be seen for all three time periods these data show Texas to have among the lowest rates of high school completion among the 50 states. In each time period, the median high school completion rate across the states was about 88%, while the completion rate for Texas was about 80%. This

Parental Opt-Out Form

Public Policy Org

new Publications francaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research
The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling

pattern indicates that the median non-completion rate across the states is about 12% while that of Texas is about 20% (about 66% worse than the median of the other states).@

Rather than take up space with all of table 7.2, which includes all 50 states, I produce here an excerpt that includes only the national averages and the relevant figures for Texas=s neighboring Mexican border states.

Table 7.2 (excerpt) High School Completion Rates of 18 Through 24 Year-olds, Not Currently Enrolled in High School or Below, by State: October 1990-92, 1993-95, 1996-98

	1990-92	1993-95	1996-98
Total National	**85.5%**	**85.8%**	**85.6%**
Arizona	81.7	83.8	77.1
California	77.3	78.7	81.2
New Mexico	84.1	82.3	78.6
Texas	80.0	79.5	80.2

Compared to its peers, Texas=s high school completion rate seems normal and steady. It isn=t rising over the years like California=s, but neither is it falling over the years like Arizona=s or New Mexico=s. It ends up higher than two of its peers= and lower than the other's.

This would have been an appropriate point in Haney=s analysis to mention the high dropout rate of Hispanic youth in the United States. The high Hispanic dropout rate is unfortunate, but it is a fact, and a well-known one at that. With its large population of Hispanics, other minorities, and immigrants, Texas should hardly be expected to have the same high school completion rate as Nebraska.

Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

3. Grade 9 Retention and High School Completion

AStates with the higher rates of grade 9 retention tend to have lower rates of high school completion... Interestingly, Texas with a grade 9 retention rate of 17.8% has a slightly lower high school completion rate (80.2%) than we would expect given the overall pattern among the states...Obviously, such a correlation between two variables, in this case, higher rates of grade 9 retention associated with lower rates of high school completion, does not prove causation, but such a relationship certainly tends to confirm the finding from previous research that grade retention in secondary school leads to higher rates of students dropping out of school before high school graduation @

Haney writes about grade 9 as is it was a magic grade, much, much more important than all the others. It just may be that he harps on grade 9 because it is there that Texas retains a high proportion of students (17%) and he wants to make a point that Texas is onerous in flunking students (and that retaining a student scars him for life and makes him want to drop out of school). Haney has no tolerance for retention by any rationale. Whether a student studies or not, whether a student learns anything or not, whether a student shows up at school or not, all students should be given high school diplomas Aon time,@ regardless. To do any less is cruel.

Even at grade 9, Texas is not the retention champ, however; New York, Mississippi, and D.C., among the states in the table below, have higher grade 9 retention rates. Pick any other grade and Texas=s retention rate is relatively low and, thus, not onerous enough for Haney=s story. So, let=s take Haney=s bait and look at grade 9.

Table 7.3Grade 9 Retention and High School Completion in the States

State	Year	Grade 9 Retention Rate	High school completion rate18-24 year-olds, 1996-98
Alabama	1996-97	12.6%	84.2%
Arizona	1996-97	7.0	77.1
District of Columbia	1996-97	18.7	84.9
Florida	1996-97	14.3	83.6
Georgia	1996-97	13.1	84.8
Kentucky	1995-96	10.7	85.2
Maryland	1996-97	10.3	94.5
Massachusetts	1995-96	6.3	90.6
Michigan	1995-96	4.8	91.0

BEST COPY AVAILABLE

Mississippi	1996-97	19.7	82.0
New York	1996-97	19.5	84.7
North Carolina	1996-97	15.8	85.2
Ohio	1996-97	11.4	89.4
Tennessee	1996-97	13.4	86.9
Texas	1995-96	17.8	80.2
Vermont	1996-97	4.8	93.6
Virginia	1995-96	13.2	85.9
Wisconsin	1996-97	8.5	90.8

Sources: Heubert & Hauser (1999) Table 6.1; Kaufman et al. (1999), Table 5.

In table 7.3, Haney attempts to show that Texas holds more students back and so more students drop out. But, I=d like to compare Texas to its peer states B other states with high minority, particularly Hispanic, and immigrant populations. Unfortunately, New Mexico and California are not in Haney=s table, so I settle for Arizona and three other jurisdictions that seem second best.

State	Year	Grade 9 Retention Rate	High school completion rate18-24 year-olds, 1996-98
Arizona	1996-97	7.0	77.1
District of Columbia	1996-97	18.7	84.9
Florida	1996-97	14.3	83.6
Texas	1995-96	17.8	80.2
New York	1996-97	19.5	84.7

Notice that Texas does not look remarkably different from its almost-peers. Arizona, which had no high-stakes testing program in the 1990s has a lower high school completion rate (ergo, higher dropout rate). Moreover, its grade 9 retention rate is about two-fifths the size of Texas=s and it still has a lower high school completion rate, contrary to Haney=s rule. The other, almost-peer states have similar grade 9 retention rates, but not remarkably higher high school completion rates. Texas does not stand out as Haney wants. Accounting for its demographics, Texas seems much like the other states.

4. Percentage of 16-19 year-olds Out of School and out of Work and Dropouts

(again)

A .. the Casey Foundation's 2000 KIDS Count on-line data base. I was alerted to this source by Hauser (1997), who ..mentions that KIDS Count project as using CPS [Current Population Survey] data in an unusual way to try to obtain relatively current evidence on dropouts across the states. Specifically, this project has compiled from CPS data three-year rolling average estimates from 1985 to 1997 of the percentage of teens ages 16-19 who are dropouts and the percentage of teens not attending school and not working.

ASuffice it to say that: 1) according to both indicators of youth welfare, between 1985 and 1997, Texas had one of the poorer records among the states, consistently showing more than 10% of teens ages 16-19 as dropouts and more than 10% of teens not attending school and not working, and 2) if one examines the standing of Texas on these two indicators relative to those of other states, conditions in Texas seemed to have worsened in the early 1990s after implementation of TAAS @

I went to the Kids Count web site of the Annie B. Casey Foundation and retrieved the relevant data. I list all the states which had a statistic of 10 percent or higher.

State in 1990	Percent of 16-19 year-olds not attending school and not working	State in 1997
Hawaii		
Idaho		Alabama
Indiana		Hawaii
Maryland		Mississippi
Michigan		Nevada
New Mexico	10	New York
North Carolina		
South Carolina		
Washington		
Alaska		
California		Alaska
Florida		Arizona

BEST COPY AVAILABLE

Illinois Missouri Texas	11	Oregon Rhode Island Texas West Virginia
Georgia Nevada Oklahoma	12	Arkansas Kentucky
Alabama Arizona Arkansas Louisiana Mississippi Tennessee	13	Louisiana Tennessee
Kentucky	14	New Mexico
West Virginia	16	

Haney is using the number of 16 to 19 year-olds who are neither in school or at work (in the aboveground economy, anyway) as a proxy measure for dropouts, even though all 19-year-olds and most 18-year-olds should be out of school if they had arrived at graduation Aon time,@ to borrow one of Haney=s favorite phrases. In other words, many of these young adults/old teenagers could simply be unemployed or working off the tax rolls.

Let=s check out Haney=s observation about Texas=s relative standing on this statistic, anyway. He claims that Texas ranks very low and got lower during the 1990s TAAS era. Notice that Texas=s percentage did not change at all in the 1990s. Is its rank very low? Not much by comparison to its peers B other Mexican border states (in orange) and other Southern states (in lilac). In 1990, three peers rank above, two are the same, and seven rank below. In 1997, two peer states rank above, one is the same, and four rank below. There are, of course, some peer states not included in the table because they have percentages less than ten (one in 1990 and six in 1997). There does seem to be a drift upwards in the table between 1990 and 1997 (i.e., toward lower percentages), but it=s certainly not uniform or overwhelming.

BEST COPY AVAILABLE

I also retrieved from the Kids Count web site the dropout data for which Haney chastises Texas. Here, I list all the states which had a statistic of 12 percent or higher.

State in 1990	Percent of 16-19 year-olds not attending school and not working	State in 1997
Georgia Kentucky Oklahoma West Virginia	12	Arkansas Florida Georgia North Carolina Rhode Island
California Florida Louisiana Tennessee Texas	13	Oregon Tennessee **Texas**
North Carolina	14	New Mexico
Alabama Arizona Nevada	15	Arizona
	17	Nevada

Again, Texas=s percentage did not change during the TAAS era, as Haney says it did, and Texas is surrounded by peer states both years, hardly near the bottom of the rankings

BEST COPY AVAILABLE

alone.

5. Immigration during the TAAS era

Alf in fact there is a net out-migration, the dropout estimates just summarized may be too high. If there is a net in-migration into Texas, the estimates are low. A...demographers conclude that between 1990 and 1995, migration into the state of Texas from other states and foreign countries increased relative to what it had been in the 1980s. They suggest that annual rates of net migration into Texas have been on the order or 1-2% in the 15 years preceding 1995. The authors do not provide direct estimates of the age distribution of immigrants into Texas, but the overall implication of their results is clear.@

This verbiage comes from a place in Haney=s report where he wants to goose the estimates (of failures, of dropouts, of retentions, of anything that he thinks will make Texas look bad). He repeats the point that youth immigration, if it could be counted properly, would certainly boost the estimates of many bad things.

If that is true where he wants it to be true, it is also true where he doesn=t want it to be true, as in the cases above, for example. The fact that Texas=s percentage of 16-19 year-olds not in school and not working stayed the same between 1990 and 1997 is good because, if you consider the number of new immigrants in the 1997 population, the proportion of those not in school or not working from the native population must have declined. Moreover, if you adjusted the high school completion rate to exclude all the immigrants who arrived during the TAAS years, the rate would have risen during the TAAS years. Finally, without the immigrants, the grade 9 retention rate would have fallen and NAEP scores for Texas would have risen even more than they did.

One of the slimier of the many slimy Aevidence gathering@ techniques of testing opponents is their use of poor and minority students. They claim to be concerned about them and interested in protecting them. But, the future they would make for them is the same as their past B years of social promotion, a diploma Aon time@ if they can stand the boredom long enough, and a life of poverty and ignorance without any practical skills.

Richard P. Phelps is the author of *Test Basher Arithmetic*, *Test Basher Benefit-Cost Analysis*, and *Why Testing Experts Hate Testing*.

http://www.edweek.org/ew/1998/26phelps.h17

http://www.edexcellence.net/issuespl/subject/standar/testbash.html

http://www.edexcellence.net/library/phelps.htm

For past articles in this series
Test Bashing Series

BEST COPY AVAILABLE

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing
Test Bashing Series
Intellectual Property and Digital Information in Higher Education
LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us
Advertising
Archives
DailyNews
Past Five Days
Book Reviews
College Aid
Book Reviews
College Aid

Test Bashing, Part 11

Walt Haney's Texas Obsession, Part 4: Dreams of Dropouts

Richard P. Phelps

Walter Haney is a zealot. He strongly believes that his views on student testing policy are best and he is determined to turn the rest of us toward his way of thinking. He doesn't much care how he does the convincing; he'll say anything.

He uses dropout statistics in Texas, and makes up some of his own, to demonstrate that the evil, devious Texas Education Agency (TEA) is cooking the stat books and that the dreaded, hated Texas Assessment of Academic Skills is, contrary to the pronouncements of the untrustworthy TEA, actually pushing more students to drop out of school.

Everybody who knows anything about dropout statistics knows that they: are very complicated; come in a wide variety of forms and magnitudes, none of which is perfect; and have low reliability by comparison with other education statistics. Walter Haney either doesn't realize this, or he realizes it full well and he's just flat-out lying to us.

Educational statisticians, such as those at the National Center for Education Statistics, have had many discussions over the past couple of decades about the many problems inherent in dropout data. But, in the end, they did agree on a common set of definitions as to what dropouts are in statistical terms. Haney completely ignores these definitions and comes up with his own. Indeed, he comes up with more than one new Haney definition.

Haney would have you believe that Texas's proportion of student dropouts has increased markedly since the introduction of the TAAS. It hasn't. He would have you believe that Texas's student dropout rate is now much higher than those in other states. Actually, it is lower. Haney would have you believe that dropout and high school completion statistics from the U.S. Education Department and the U.S. Census Bureau are all wrong, and that he's figured out what the real dropout rate is in Texas. He hasn't.

Turns out, Texas actually has rather low rates of dropout, low by national standards, and lower than demographically similar states.

Some of Haney's "mistakes" in interpreting dropout and high school completion data may, indeed, be the result of an ignorance of the statistics. In some cases, however, it is pretty difficult to believe that he couldn't know that he's being deceitful.

Haney's own definition of "dropout"

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
new
Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

Take this sentence from his report's abstract: "Only 50% of minority students in Texas have been progressing from grade 9 to high school graduation since the initiation of the TAAS testing program." ...and this section from his report: "...these results lead me to conclude that since the implementation of the TAAS high school graduation test in 1991, 22-25% of White students and 35-40% of Black and Hispanic students, have not been persisting from grade 6 to regular high school graduation six years later. Overall, during the 1990s the dropout rate in Texas schools was about 30%. As appalling as this result appears..."

In the first sentence from the abstract, Haney is counting any minority student from grade 9 who does not graduate three years later as a dropout without, of course, explaining that to the reader. By "not progressing to high school graduation since the initiation of the TAAS testing program," he really means not progressing to high school graduation "on time," that is, exactly three years later. Some of those students are retained in grade somewhere in between the start of grade 9 and the end of grade 12, of course. But, they become dropouts by one of Haney's definitions. As Haney makes plain in several parts of his report, not graduating "on time" is, in his opinion, an "appalling" scandal. He believes that no student should be restrained in any way from graduating "on time," regardless the reason.

(The overwhelming majority of U.S. citizens, of course, disagree with Haney. They believe that students should graduate after they've met generally agreed-upon academic standards, and not before. Moreover, most citizens believe that students should be able to demonstrate mastery of those standards before being allowed to graduate.)

His way of counting dropouts becomes more explicit in the second quote in the paragraph above where Haney uses the phrase "regular high school graduation," by which he means graduation "on time." With his "on time" graduation percentages, he can increase the number of alleged dropouts by counting as dropouts those students retained one year who graduate in their 13th year of school.

(I don't have time to figure out why his minority dropout rate was 50% in his first quote and 35-40% in his second.)

Here's another example where the ruse is even clearer:

"If we adopt the common sense definition that a dropout is a student who leaves school without graduating from high school, analyses of data... tell a reasonably clear story of what has happened in Texas over the last two decades...the percentage of Black and Hispanic students who progressed from grade 6 to graduation six years later hovered around 65%. For White students, the corresponding percentage started at about 80% and gradually declined to about 75% in 1990."

Haney tells us, first, that a dropout is a student who leaves school without graduating and then, second, tells us that he is counting students who progressed from grade 6 to graduation 6 years later. Ergo, those who graduate 7 years later, in their 13th year of school, are being counted by Haney as dropouts.

One final quote to hammer home the point:

BEST COPY AVAILABLE

Parental Opt-Out Form

Public Policy Org

Publications francaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

" ..to be absolutely clear (and to avoid getting into semantic arguments about the meaning of the term "dropout"), I readily acknowledge that what the cohort progression analyses show is the extent of the problem in Texas of students failing to persist in school through to high school graduation----regardless whether it is caused by students having to repeat grade 9, failing to pass the exit level version of TAAS, officially "dropping out," opting out of regular high school programs to enter GED preparation classes, or some combination of these circumstances."

To Haney, the "official" dropout rate is just one of several components of his, better dropout rate. Some of the world's foremost education statisticians have been wrestling with the issue of how to count and define dropouts but, who cares? Walt Haney has come along and found a better way. Or, perhaps, he wants the reader to have in mind the "official" definition while he actually uses his "absolutely clear" definition to calculate (larger) dropout numbers for Texas.

The Texas Education Agency Must be Lying to Us, Again

Haney lambasts the TEA early and often in his report. He paints a portrait of institutionalized dishonesty and rampant incompetence. Their estimates of dropouts seem low to him, for example, so he accuses them of cooking the stat books:

"It is clear that the TEA has been playing a Texas-sized shell game on the matter of counting dropouts. Every source of evidence other than the TEA (including IDRA, NCES, the Casey Foundation's KIDS Count data, Fassold's analyses and my own) shows Texas as having one of the worst dropout rates among the states. (Recall that even the Texas State Auditor's Office estimated that the 1994 dropout numbers reported by the TEA likely covered only half of the actual number of dropouts.)"

Haney doesn't tell you that ALL states underestimate their dropout numbers. It is a function of how state education agencies collect dropout data. They cannot conduct a general population survey like the Census Bureau does with its Current Population Survey, the source of federal statistics on dropout rates and high school completion. States are obligated by law to collect the data that the local school districts give them, including that on dropouts, and the districts have every incentive to delay notifying the state of their dropouts. Likewise, individual schools have every incentive to delay notifying the district of their dropouts. At best, dropout numbers arrive at the state level very late; at worst, schools and districts invent a variety of means of deliberately hiding those students who initially enrolled but no longer show up for class.

Districts and schools do this because, in most states, they are allocated state education funds based on the number of students they have in attendance. Some states do better jobs than others at monitoring these numbers, but vigilant monitoring does not come cheap. It requires frequent state inspection of classroom attendance and cross checking names on enrollment and attendance collection forms.

Nonetheless, even though Texas's official state dropout numbers almost certainly underestimate the true magnitude of dropouts, if the data collection is consistent over the years, trends in the state averages should still be informative. The TEA shows dropout numbers trending downward during the 1990s, during the TAAS era, and asserts that the TAAS may actually be decreasing the number of dropouts and certainly cannot be increasing it.

BEST COPY AVAILABLE

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

The TEA could be right. At least Haney's complaints to the contrary are no threat to the claim. Haney argues that since Texas's dropout numbers are underestimates, the trend must be incorrect. His reasoning is not valid. If the numbers are underestimated in a consistent way over the years, they can provide significant evidence of a trend toward fewer dropouts, even if the absolute magnitude of the numbers is off.

Evaluating Texas on Accurate and Reliable Dropout Measures

There's a table in the NCES publication *Dropout Rates in the United States: 1998* that Haney does not cite, even though he cites other sections of the report. It lists all 50 states according to their "event" dropout rates, one of the accepted, standard ways of measuring dropouts. An "event" rate is the proportion of students who leave school each year without completing a high school program. Specifically, the event rate in this publication measured the number of 15 to 24 year-olds who were enrolled in high school in October 1997, but were not enrolled a year later and had not completed high school. To get around the problems inherent in administrative dropout numbers, this information is captured in the U.S. Census Bureau's October Current Population Survey (the October collection is larger than other months').

How does Texas look? It has a 3.6 percent event dropout rate, lower than the U.S. average, despite its demographic disadvantages. Its Mexican Border neighbors, Arizona and New Mexico, have rates of 10.0 and 7.5 (California has no rate listed). The average rate for the South is 5.1.

Texas has the sixth lowest event dropout rate in the United States in 1996-97, the most recent year measured. The official statistics, calculated by some of the world's foremost statistical experts, buttress the TEA's claims about Texas's low dropout rate.

Especially given Texas's demographic disadvantages (Hispanic dropout rates are *triple* the average), its school officials deserve high praise. Instead, after doing such a wonderful, commendable job, they must endure the cynical attacks of test bashers like Haney, and the journalists who blindly believe such "research."

Haney claims, "Every source of evidence other than the TEA (including IDRA, NCES, the Casey Foundation's KIDS Count data, Fassold's analyses and my own) shows Texas as having one of the worst dropout rates among the states." Shameless.

The same NCES publication shows a table of state-by-state high school completion rates. Texas doesn't look as good here compared to the national average. Haney makes a big deal out of this table in his report, even while he completely ignores the dropout table. .

That Magic Grade 9 and its Relation to High School Completion

Haney writes about grade 9 as is it was a magic grade, much, much more important than all the others. He draws a scatterplot that shows an almost perfect (inverse) correlation between grade 9 retention rates and high school completion. Grade 9, he wants us to believe, is very special, the gateway grade to high school graduation. Believe it if you like.

BEST COPY AVAILABLE

"The grade 9 retention rates in Texas are far in excess of national trends. The recent report of the National Research Council (NRC) also shows Texas to have among the highest grade 9 retention rates for 1992 to 1996 among the states for which such data are available (Heubert & Hauser, 1999, Table 6-1).

"States with the higher rates of grade 9 retention tend to have lower rates of high school completion....Interestingly, Texas with a grade 9 retention rate of 17.8% has a slightly lower high school completion rate (80.2%) than we would expect given the overall pattern among the states....Obviously, such a correlation between two variables, in this case, higher rates of grade 9 retention associated with lower rates of high school completion, does not prove causation, but such a relationship certainly tends to confirm the finding from previous research that grade retention in secondary school leads to higher rates of students dropping out of school before high school graduation."

It just may be that Haney harps on grade 9 because it is in that grade that Texas retains a high proportion of students (17%), and he want to make Texas appear onerous in flunking students. In fact, looking at all grades, and not just grade 9, Texas has relatively low retention rates.

Even at grade 9, Texas is not the retention champ, however; New York, Mississippi, and D.C. have higher grade 9 retention rates. Comparing Texas to its peer states - other states with high minority, particularly Hispanic, and immigrant populations, Texas does not look remarkably different. Arizona, which had no high-stakes testing program in the 1990s has a lower high school completion rate (ergo, higher dropout rate). Moreover, its grade 9 retention rate is about two-fifths the size of Texas's and it still has a lower high school completion rate, contrary to Haney's rule. Texas does not stand out as Haney wants. Accounting for its demographics, Texas seems much like the other states.

Looking beyond grade 9, Texas looks even better. Here's my partial reproduction of one of Haney's favorite tables, the state-by-state retention rate table (table 6.1) from the Heubert and Hauser book. Starting on the book's p.138 are several pages of state-level and grade-level retention rates, comprised by two researchers who share Haney's slant on testing issues almost perfectly. I'm suspect of its accuracy. But, again, I'm going to take Haney's bait and go along with him. Not all U.S. states are included in the table and, even for those that are included, data are often missing for certain grades. Fortunately, most Southern states are included, with fairly recent data. I want to compare Texas, again, to its peers.

BEST COPY AVAILABLE

	grade									gr. 1--9	gr. 1--9
state	1	2	3	4	5	6	7	8	9	sum	rank
virginia	4.3	2.4	1.9	1.6	1.1	3.6	5.3	6.0	13.2	39.4	4
texas	**5.9**	**2.6**	**1.5**	**1.0**	**0.8**	**1.7**	**2.9**	**2.1**	**17.8**	**36.3**	**7**
tennessee	5.5	2.5	1.8	1.2	1.4	2.7	7.2	5.7	13.4	41.4	3
n carolina	5.7	3.1	2.5	1.4	1.0	2.6	3.4	2.8	15.8	38.3	5
mississippi	11.9	6.6	5.4	6.1	6.6	7.7	15.6	12.9	19.7	92.5	1
georgia	4.0	2.4	1.7	1.3	1.1	2.1	2.5	2.1	12.4	29.6	8
florida	4.1	2.2	1.5	1.0	0.7	4.4	4.9	4.0	14.3	37.1	6
arizona	2.2	1.0	0.7	0.5	0.5	1.1	2.7	2.3	7.0	18.0	9
alabama	8.5	3.3	2.5	2.1	2.0	2.9	6.1	4.4	12.6	44.4	2
average	5.8	2.9	2.2	1.8	1.7	3.2	5.6	4.7	14.0	41.9	N=9

Does Texas retain students at a rate much higher than other states, as Haney claims? Absolutely not. Indeed, grade 9 is the only grade where Texas's retention rate is above average (except for grade 1's 0.1 difference). No wonder Haney picked grade 9. Any other grade, 2 through 8, and Texas looks like a wimpy social promotion state, the kind that Haney reveres.

Texas comes in below the average sum for grades 1 through 9, ranking 7th in cumulative retention rate out of only 9 states. Texas ranks very low on retention.

A look at the upper grades shows much the same picture (see table below). Texas ranks below average on retention for grades 10 and 12, and slightly above average for grade 11. Its cumulative retention rate for all grades 1 through 12 is well below average, ranking 6th out of only 9 states in retention.

	grade			gr.1-12	gr.1-12
state	10	11	12	sum	rank
virginia	8.4	6.2	6.4	60.4	4
texas	**7.9**	**5.5**	**4.2**	**53.9**	**6**
tennessee	9.5	7.0	5.8	63.7	2
n carolina	6.8	2.1	4.7	51.9	7
mississippi	12.8	7.7	5.2	118.2	1
georgia	8.7	5.4	3.5	47.2	8
florida	8.6	5.7	5.0	56.4	5
arizona	5.0	3.1	10.2	36.3	9
alabama	6.7	5.2	5.1	61.4	3
	8.3	5.3	5.6	61.0	N=9

Grade 9 Enrollments and Graduates - "Persistance" Across Grades

Haney also makes a big deal out of comparing grade 9 enrollments to numbers of

graduates and, for that matter, the enrollments in other grades to the number of graduates, both "on time" and the kind the rest of us are familiar with. One result is the particular type of Haney-special dropout statistic that he ends up with in his report.

His words themselves are virtually breathless as he writes:

"Overall, during the 1990s the dropout rate in Texas schools was about 30%. As appalling as this result appears. ...A convergence of evidence indicates that during the 1990s, slightly less than 70% of students in Texas actually graduated from high school."

Haney claims that this 30% is the "real" dropout rate for Texas. He also strongly implies that it is unique to Texas, that uniquely evil state that rises above all the others in ill will, dishonesty, incompetence, and bad education policy. (Don't tell him that the TAAS is a lot like several other state testing programs.)

He gets the 30% by comparing enrollments in a middle- school level grade X, like grade 9 or 8 or below, to the number of graduates (12 - X) years later. To check his claims, I use his own numbers, from the spreadsheet appendix to his report. Dividing the number of graduates in Texas for the 1998-99 school year by the number of 8th graders in 1995-96, I get 0.72, almost exactly what Haney led me to expect. If I do the same for grade 9, I get 0.61, but, remember, there's that big retention bulge in grade 9 that distorts comparisons (remember, that's why Haney likes it).

Mind you, this is NOT a dropout rate, as Haney claims. I don't have time to get into the nuances of the definitions now, maybe later.

Right now, I'm curious to see HOW MUCH higher Texas is on this measure than other states. Haney writes that Texas is a lot higher in dropout rate and a lot lower in persistence to graduation than other states. But, he's comparing apples and oranges; his special measure for Texas with genuine dropout measures for other states. How do other states stack up on his measure?

It so happens, a lot like Texas. Texas is not unique as Haney would have us believe.

I used data from the U.S. Education Department's *Digest of Education Statistics*, comparing graduates in 1997-98 to 8th graders in 1993-94 and 9th graders in 1994-5, in nine states demographically similar to Texas. The persistence rates below are calculated by dividing graduates by each grade's enrollments. These constitute measures of how may of the 8th or 9th-grade students were still in school X years later when they were supposed to be.

	us	**tx**	ca	az	nm	la	ms	al	ar	tn	nc
8th gr. Enrol (000) 93-94			389	55	25	58	40	59	36	66	86
Grads (000) 97-98			269	36	17	38	24	38	26	44	59
persistence rate	0.71	**0.72**	0.69	0.65	0.68	0.66	0.60	0.64	0.72	0.67	0.69
9th gr. Enrol (000) 94-95			438	59	29	69	43	65	37	74	101
persistence rate	0.64	**0.61**	0.61	0.61	0.59	0.55	0.56	0.58	0.70	0.59	0.58

BEST COPY AVAILABLE

Far from standing out from the crowd with the unusually low persistence rate Haney claimed we would find, Texas has a higher-than-the-national-average persistence rate and a rate higher than those for all the comparison states, save one, for the grade 8 comparison.

For the grade 9 comparison, Texas is hobbled by the grade 9 retention bulge and, not surprising, its persistence rate of 0.61 slides under the national average this time. However, compared to its peer states, Texas is above average, higher than six out of nine.

But, Texas is Really, Really Low

Let's look again and further into what I referred to as Haney's shameless claim. Again, he wrote: "Every source of evidence other than the TEA (including IDRA, NCES, the Casey Foundation's KIDS Count data, Fassold's analyses and my own) shows Texas as having one of the worst dropout rates among the states."

I don't have the time or resources to investigate the sources from organizations and individuals who are affiliated with or otherwise share Haney's true objectives (i.e., IDRA and Fassold) but I can check Haney's claim that NCES and KIDS Count data show "Texas as having one of the worst dropout rates among the states."

As I asserted in last week's commentary, Haney brings up the topic of adjusting statistics for demographic considerations whenever it serves his goal of lambasting the TEA, and he ignores demographic considerations whenever it does not serve his goal of lambasting the TEA. Throughout his report, he criticizes Texas for being a low-ranking state on this or that statistic. He compares Texas to all the other U.S. states - Minnesota, Vermont, Oregon, Wyoming. On statistics that are likely to be affected by a state's demographic profile -- statistics like dropout rates and high school completion - does it make much sense to compare Texas to Minnesota, Vermont, Oregon, and Wyoming? Of course not.

If one compares Texas to its demographic peers, the other Southern or Mexican Border states, Texas comes out above average. Probably the most pertinent demographic factors as regards effect on dropout rate are the percentages poor and minority of a student population in a state. It is, indeed, unfortunate, that Black and, particularly, Hispanic students tend to have higher dropout rates, but they do. And, Texas has a much larger Black population than does Wyoming. Texas has the second largest population of Hispanic students in the country, second only to California. Moreover, Texas's minority population tends to be poor.

NCES data on high school completion

Let's look at the NCES dropout/high school completion data, first. Rather than take up space with all of Haney's table 7.2, which includes all 50 states, I produce here an excerpt that includes only the national averages and the relevant figures for Texas' neighboring Mexican border states.

Table 7.2 (excerpt)
High School Completion Rates of 18 Through 24 Year-olds,

BEST COPY AVAILABLE

**Not Currently Enrolled in High School or Below,
by State: October 1990-92, 1993-95 and 1996-98**

	1990-92	1993-95	1996-98
Total National	**85.5%**	**85.8%**	**85.6%**
Arizona	81.7	83.8	77.1
California	77.3	78.7	81.2
New Mexico	84.1	82.3	78.6
Texas	80.0	79.5	80.2

Compared to its Mexican Border peers, Texas's high school completion rate seems normal and steady, just like the country's as a whole. It isn't rising over the years like California's, but neither is it falling over the years like Arizona's or New Mexico's. It ends up higher than two of its peers' and lower than the others'.

The results are similar in comparing Texas to its Southern States' peers. Here, Texas does have the lowest high school graduation rates, but they are not appreciably lower than those in Louisiana and Mississippi, its neighbors, or in Florida, the only other state in the list with a sizeable immigrant population.

Table 7.2 (excerpt 2)
High School Completion Rates of 18 Through 24 Year-olds,
Not Currently Enrolled in High School or Below,

	1990-92	1993-95	1996-98
Total National	85.5%	85.8%	85.6%
Alabama	83.9	83.6	84.2
Arkansas	87.5	88.3	84.5
Florida	84.1	80.6	83.6
Georgia	85.1	80.3	84.8
Louisiana	83.9	80.1	81.6
Mississippi	85.4	83.9	82.0
No. Carolina	83.0	85.5	85.2
So. Carolina	85.0	87.8	87.6
Tennessee	75.7	84.5	86.9
Texas	80.0	79.5	80.2

Source: Kaufman, P., Kwon, J., Klein, S. and Chapman, C. (1999). Dropout rates in the United States: 1998. (NCES 2000-022). Wash., D.C.: National Center for Education Statistics, p. 20.

KIDS Count data

Now, let's look at the Annie E. Casey Foundation's KIDS Count data. Since I took on this issue last week, I won't include the rather long tables comparing Texas to its peers. The reader can view them in section 5 of:

http://www.educationnews.org/test_bashing_part_10.htm

Here's the relevant criticism from Haney:

"Texas had one of the poorer records among the states, consistently showing more than 10% of teens ages 16-19 as dropouts and more than 10% of teens not attending school and not working; and if one examines the standing of Texas on these two indicators relative to those of other states, conditions in Texas seemed to have worsened in the early 1990s after implementation of TAAS."

The first KIDS Count measure is the number of 16 to 19 year-olds who are neither in school or at work (in the aboveground economy, anyway) -- a proxy measure for dropouts. All 19-year-olds and most 18-year-olds should be out of school if they had arrived at graduation "on time," however, to borrow one of Haney's favorite phrases. In other words, many of these young adults/old teenagers could simply be unemployed or working off the tax rolls.

Haney claims that Texas ranked very low and got lower during the 1990s TAAS era. The KIDS Count data, however, show that Texas's percentage did not change at all in the 1990s. Was its rank very low? Not much by comparison to its peers - other Mexican border states and other Southern states. In 1990, three peers rank above, two are the same, and seven rank below. In 1997, two peer states rank above, one is the same, and four rank below. There are, of course, some peer states not included in the table because they have percentages less than ten (one in 1990 and six in 1997). There does seem to be a drift upwards in the table between 1990 and 1997 (i.e., toward lower percentages), but it's certainly not uniform or overwhelming.

I also retrieved from the Kids Count web site the *dropout* data for which Haney also chastises Texas. Here, I compare all the states which had a statistic of 12 percent or higher. Again, Texas's percentage did not change during the TAAS era, as Haney says it did, and Texas is surrounded by peer states both years, hardly near the bottom of the rankings alone.

In fair comparisons of demographic peers, then, Haney's derision of Texas as a "bottom state" is just wrong.

Genuine, reliable, and accurate data on dropouts shows Texas to have a relatively low rate. Because this fact does not square with Walt Haney's preconceived bias, he makes up his own measures of dropout rate. His measures are rickety constructions that can't hold up against a couple of hours of casual scrutiny. His alleged evidence is worthless.

If the TAAS deserves severe criticism, it will have to come from someone other than Walt Haney. His criticisms might keep people misinformed through the election season, but they won't last much longer than that.

Richard P. Phelps is the author of *Test Basher Arithmetic*, *Test Basher Benefit-Cost*

BEST COPY AVAILABLE

Analysis, and *Why Testing Experts Hate Testing*:

http://www.edweek.org/ew/1998/26phelps.h17

http://www.edexcellence.net/issuespl/subject/standar/testbash.html

http://www.edexcellence.net/library/phelps.htm

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

Shop Here for Books
amazon.com

CAPAlert for Parents & Grandparents

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 12

Education Week's Anti-Testing Bias: It May Start at the Top

Richard P. Phelps

Guess who wrote the words below:

"Of all the school reform ideas circulating today, high-stakes student testing has to be about the worst. Unfortunately, instead of slowing down and reassessing the use of these tests, more and more states seem determined to make them the engine of their school-improvement strategies. Apparently they would rather risk a public backlash than be perceived as 'backsliding.'

"Those who push for high-stakes testing contend that the threat of severe penalties pressures students and teachers to improve performance. Schools are doing poorly, they imply, because students and teachers aren't working hard enough. This argument brushes aside decades of research that links poor performance to the way schools are organized, operated, governed, and funded. It also ignores the impact that poverty and discrimination have on student performance."

The two paragraphs above are filled with the standard premise and argument of testing opponents on education school faculty and in some well-known extremist advocacy groups. There's a call for more "research" before standardized testing is administered for high-stakes and, presumably, the research will be conducted by testing's opponents, with predictable results. It is assumed that the alleged "backlash" against testing is derived from widespread and new public opposition rather than the activities of a few small pressure groups that have long opposed testing. The writer adds the now-all-too-familiar charge that supporters see high-stakes testing as a "silver bullet" cum "quick fix" cum "cheap fix" without identifying any testing supporters who are really so narrow-minded, probably because there are few or none.

Finally, the writer cites "the research" which purports to show that poor student performance has nothing to do with what and how students are taught. If even half of the research conducted by the radical constructivists and radical egalitarians (hereafter, "utopians," for short) whom he trusts was any good, I might agree with him.

There are, generally, two types of research: that which looks at data and evidence first and then derives conclusions based on the evidence, and that which starts with the conclusions first and then molds the data and evidence to fit the conclusions. Much, if not most, of the anti-testing research conducted by the utopians is of the second type. It is not the type of objective research that one finds in mature, self-secure professions like physics or

BEST COPY AVAILABLE

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

economics. It is advocacy research, the type one sees in political advertising at campaign time.

One type of anti-testing advocacy research incorporates the lying-with-statistics method - numbers are left out if they don't serve the conclusion, data are altered, costs are made up, benefits are ignored, numbers are labeled with misleading descriptions, definitions are changed surreptitiously, supportive research is cited that does not contain the evidence claimed, and solid research that does not support the preferred conclusion is ignored.

More common than the lying-with-statistics studies, however, are those that use semantic distortions, with only illusions of data.. By "semantic distortion," I mean: lots of scientific-sounding language used without any real science behind it; lots of research-like language used without any valid research behind it; educational practices that testing opponents like are given euphemistic labels and referred to with attractive language, without descriptions of what the practices entail in practical terms; and educational practices that testing opponents do not like are referred to with unappealing language, without descriptions of what the practices entail in practical terms.

Most of the education "research" that attacks the use of standardized tests is more pseudo-science than science, but it is usually very strongly worded, uses lots of terminology that sounds sorta scientific, is written by folks with academic credentials from (often) prestigious universities, and these folks have developed an expertise in phraseology that works (to persuade naive (or ideologically sympathetic?) journalists, for example).

As they say, you can get data to say anything if you torture them long enough. Likewise, you can get words to mean anything if you torture them long enough. But, hey, it works. It may work well enough to convince the chairperson of *Teacher Magazine* and former editor of its sister publication, *Education Week*, for example. He authored the quotes listed above and below in *Teacher Magazine*, October 1.

Here are some more accusations of high-stakes testing made by the high official of "Education's Newspaper of Record," along with my responses:

(1) "States are using standardized test scores as the single measure to determine whether a student passes or graduates, which is stupid and unfair."

No, they are not. Students are given several to many chances to pass these high school exit exams. Moreover, these exams are typically set at a middle-school level of difficulty. It is very sad to see *Teacher Magazine* and *Education Week* supporting this, perhaps testing opponents' most disingenuous charge.

Exit exam requirements are no different from any other high school completion requirement. If a state requires that students complete four levels of English to graduate, no student can graduate without passing four levels of English. If a student fails a senior-level English course, she does not graduate. And, ultimately, *at the margin*, she can fail simply for not passing an end-of-semester exam in any level of English, by just one question.

What's true for English is true for any other graduation requirement, from passing grades in four levels of Physical Education courses to completion of minimum amounts of community service time in some states.

BEST COPY AVAILABLE

Parental Opt-Out Form

Public Policy Org

Publications françaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

community service time in some states.

(2) The editor accuses testing supporters of opposing "opportunity-to-learn standards" only because it would have cost more money.

In fact, many testing supporters also supported the implementation of opportunity-to-learn standards and those that didn't usually didn't for reasons other than cost (e.g., the ideological distortions they thought would be inevitable in the implementation of some of the standards).

(3) "Many middle and high school students--especially those in the most disadvantaged neighborhoods--do not read well and have not had an opportunity to learn the material required to meet the standards set for them....the teachers generally assigned to the most at-risk students are those least prepared for the task [to teach to standards]."

Whose fault is it that students do not have an opportunity to learn required material? Whose fault is it that teachers are unprepared to do their jobs? Are we to wait until the last of the incompetently-run schools or districts in America gets its act together before we implement accountability programs? We would then, of course, be waiting forever.

(4) The editor describes multiple-choice test items as "primitive."

Contemporary multiple-choice tests are the most technically-advanced and reliable (i.e., fair) type of standardized test in the world. It is open-ended tests that are "old-tech." Mind you, that doesn't make open-ended tests all bad. Both types of tests have advantages and disadvantages, but the alleged disadvantages of multiple-choice tests have been way overblown of late and their advantages largely ignored.

(5) "Many states use 'off-the-shelf' standardized tests that are not aligned with the standards and curricula they've installed in schools. In some cases, high stakes are attached to norm-referenced tests, which rank students according to the performances of their peers, not to any academic benchmark. How logical is that?"

I agree that it is not fair to make high-stakes decisions about students based on tests that are not matched to a required curriculum. But, if today's testing programs are not as good as they should be, the editor need look no further for blame than his own newspaper and magazine.

Any more, most of the utopians' anti-testing research doesn't even make much of an effort to hide the deceit. They've gotten away with stretching the truth for so long (with the docile participation of the media) that they now think that they can say anything they please. It takes very little effort (witness my columns of the past four weeks) to dissect the "research" to find the fudged numbers, lack of evidence, semantic distortions, and made-up facts.

It's sad that it is left to independent outsiders like me to do this work that established education research institutions should be doing. There are a thousand other activities that I would prefer occupied my weekend time than writing these commentaries. I write them because the institutions that should be conducting this kind of critical quality control on testing research, institutions like *Education Week* and the Center for Research on Evaluation, Standards, and Student Testing (CRESST), do not. Indeed, CRESST is itself

BEST COPY AVAILABLE

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

responsible for producing much of the junk research, and they do it with our tax dollars.

For its part, and with the single exception of the Commentary Editor, Sandy Reeves, *Education Week* presents junk research uncritically, completely ignores all research done on the benefits of testing, completely ignores critiques of the junk research, and sends its reporters to conferences of organizations captured by the utopians (e.g., CRESST, AERA, Board on Testing and Assessment at the NRC) while ignoring the conferences where objective education research is presented (e.g., APPAM, NCME, APA, ASA, AEA)

If *Education Week* and *Teacher Magazine* were balanced and objective in their coverage of testing issues, it could fill a void and serve a critical public need.

Most policymakers do not trust the utopians' research on testing, nor should they. By choosing to serve as an uncritical conduit for all the most dishonest and self-interested "research" of the utopians, *Education Week* and *Teacher Magazine* only help to increase the unfortunate but understandable cynical distrust toward all education research held by so many thoughtful people.

Richard P. Phelps is the author of *Test Basher Arithmetic*, *Test Basher Benefit-Cost Analysis*, and *Why Testing Experts Hate Testing*:

http://www.edweek.org/ew/1998/26phelps.h17

http://www.edexcellence.net/issuespl/subject/standar/testbash.html

http://www.edexcellence.net/library/phelps.htm

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 13

The Sad End of Objectivity at the Education Commission of the States

by Richard P. Phelps

October 24. 2000

On the surface it would appear a heartening and beneficial change. After biting the hands that have fed it very well for two decades, the taxpayer-funded student testing research center, CRESST, now says it will "help" the bipartisan Education Commission of the States (ECS) to implement state high-stakes testing programs. In the face of overwhelming public support for high-stakes tests, CRESST has consistently told its benefactors for twenty years that they shouldn't bother, as the "research" shows that high-stakes tests, multiple-choice tests, and pretty much any meaningful standardized student tests are terrible things.

Those with a cynical view of the research conducted at the Center for Research on Evaluation, Standards, and Student Testing (CRESST), like me, suspect that CRESST's motivations have little to do with any genuine research findings but, rather:

- It is in the self-interest of education professors to stonewall any effective program of accountability in our public school systems (so they can be left alone to do whatever they please instead of being held accountable for working toward improving student achievement); and
- Believers in two utopian ideologies, radical egalitarianism and radical constructivism, have taken over essential control of the institutions of education research in most fields outside the more technical ones, such as psychometrics, quantitative methodology, and education finance. Most CRESST researchers are and have been education professors.

For CRESST, this has meant a consistent production of a substantial quantity of top-quality psychometric research, conducted by some of the world's foremost, brilliant technical minds. The quality of CRESST's work takes a nose dive, however, when its researchers

Ed. Links

Ed. Org

ELA

Grammar

Hispanic Links and Resources

History

Home School

Job Postings

K-12 Science

NEW
Learning Disabilities and High-Stakes Standardized Tests

Math

Media Links

Mental Health

Texas Reading Institute

No Child Left Behind

National Reading Panel Report

Phonemic Awareness

Reading Kits

Reading Tests

Ideas Practicas Para Padres de Familia
pdf file

Special Education Resources

TAD for English/Language Arts/Reading

LA Times Special Report
Special Education a Failure on Many Fronts

venture into the field of public policy. Along side the lode of psychometric gemstones lay some of the most tawdry, self-serving, transparently-biased ideological treatises ever produced with U.S. taxpayer dollars.

By contrast, the Education Commission of the States is quite a different animal. Centrally located in Denver, its purpose is to aid elected officials in all U.S. states formulate education policy and implement education programs. ECS staffers, who are charged with informing politicians of both political parties and a wide range of political persuasions, are themselves among the best-informed observers of the education scene in the United States. They know all sides of the issues on which they specialize and they keep up with them on a daily basis.

I have not always been happy with ECS pronouncements, however, as I think ECS staffers are sometimes too trusting. Just this year, for example, they swallowed whole the ruse of anti-testing groups that a nationwide groundswell of "independent, grassroots, parents'" organizations had risen up in opposition to high-stakes testing. Some ECS staffers have also swallowed whole the full bore versions of "teaching to the test," "narrowing the curriculum," and other tenets of faith relentlessly pushed by anti-testing advocates.

In the defense of ECS, however, it is not a large organization and it has the full menu of education issues on its plate, not just testing. Indeed, there are really only one or two staffers at any given time who specialize in testing issues and programs and only several others who even know testing issues tangentially. Nonetheless, ECS staffers are kept busy by a steady stream of telephone calls from enlightened journalists who confidently rest assured that their responses will be sincere and objective. Journalists don't get that from CRESST, nor from most education school faculty, for that matter. ECS has successfully maintained an oasis of honesty in a desert of deceit.

Until now, that is. With all the work ECS staffers have to do, it will be just too tempting to let CRESST "handle" the testing issues. Moreover, even if the one or two testing experts at ECS try to hold their ground, they will be greatly outnumbered by dozens of CRESST "researchers" and their academic friends.

What kind of "information" about student testing can we expect from the new CRESST/ECS partnership? That's an easy question to answer. Just look at CRESST's past behavior.

Probably, the most important feature of CRESST research on testing policy issues is the degree to which it narrows the topic. You would think that with the many millions of dollars we taxpayers have given

Parental Opt-Out Form

Public Policy Org

Publications francaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research
The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

them over the past twenty years, CRESST reports would be expansive and wide-reaching. Just the opposite has been the case. The reports on policy issues have used very little source material, generally that to be obtained from CRESST researchers themselves and a small circle of friends.

This narrowness has two effects: it serves to promote the careers of CRESST researchers; and it masks the abundant research and evidence that might lead one to conclusions other than those favored by CRESST. While twenty years of CRESST reports have amply discussed a plethora of real and (mostly) imagined problems with high-stakes testing, for example, not a single one out of the hundreds has broached the subject of high-stakes testing's benefits. Nor have any CRESST reports broached the topic of the considerable validity, reliability, and fairness problems associated with the exclusive reliance on the traditional alternatives to standardized tests, such as student grade point averages, that CRESST favors. Anyone who thinks they will read broadly "the state of knowledge" or the "state of the art" in CRESST reports will be disappointed. They'll read only one side of a very thin slice.

As I don't have the time, and you probably don't have the patience, for an exhaustive review of CRESST reports on state testing programs and testing policy, I will limit my comments to just two prevalent CRESST themes:

1) "Politicians" and the general public have no business making testing policy or running testing programs. These tasks should be left to testing "experts" and "educators," who are familiar with the "research" on testing. CRESST terminology uses the term "educators" rather more narrowly than most of us do. Roughly, "educators" are those involved in education who share CRESST's point of view. Specifically, psychometricians with Ph.D.s in education who devote their lives to writing fair, useful, and technically sound standardized student tests are not "educators" if they work in the private sector for a test development firm. Those who work in state education agencies are not educators unless they have degrees in education. Parents are not educators. Professors who devote their study to education issues are not educators unless they are education professors with education degrees. Education professors, however, are definitely educators, even though they do not work in the K-12 system at all.

2) Standardized tests are very, very, very, expensive, much more expensive than you probably think, and they cost more than you probably want to pay.

My favorite examples of the first approach-that "non-educators" have no business being involved in testing-come from a series of CRESST

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

reports that ridicule the government of the state of Arizona and vilify its current state superintendent, Lisa Graham Keenan. The series is comprised of, at least, CRESST reports 321, 373, 380, 381, 420, 425, 468. (Yes, these are big numbers; you and I have paid for over 500 CRESST reports to date.) CRESST report number 426 has my favorite title "The Politics of Assessment: A Case Study of Policy and Political Spectacle."

Here are some excerpts from this series of CRESST reports on the "Political Culture"of Arizona:

"The dominant discourse was union-bating and educator-bashing, federal mandate- and court order-defying. Right-wing extremists often made the news, as did religious conservatives. Assessment policy could hardly be immune from this climate, particularly because of the relationship between political and pedagogical conservatism...."

"...right-wing organizations typically extol the virtues of ...phonics...and math by memorization of math facts. They repudiate the teaching of higher-order thinking, whole language and bilingual education, and other recommended approaches of progressivism and constructivism (Dewey and Vygotsky's communist leanings make these perspectives suspect)....

"These preferences have the characteristics of fixed ideologies, in that they seem founded in biblical interpretation, are immune to fair debate, and tend to demonize the opposition...."

"Noneducator interests [in Arizona] dominate policy making over educators'. The primary policy value in the state is efficiency (tax savings) rather than excellence or equity. Education was defined as an economic function long before it became so defined at the national level....Arizona is a right-to-work state, and teachers have very little say in a climate that systematically dismisses them.

"The media also play a role in [Arizona's] political culture. The two newspapers are owned by Dan Quayle's family. They express the values of efficiency and antiprofessionalism on a daily basis...They never mention an educational issue without using the term "educational establishment." With great relish, they publish the yearly results of student assessments and use these or any indicators as the source of editorial handwringing about the failure of public schools.

"In spite of Arizona being near the bottom in spending on education, health, and social programs...the legislature...passed the largest tax decrease in state history...the polluter protection bill..the 'veggie hate crimes' bill...charter school legislation [and] the governor...sought to

BEST COPY AVAILABLE

avoid the federal mandates to provide school services to immigrants and to protect endangered species and fragile ecosystems."

"Against this landscape of political culture, the organization of schooling struggles."

"[CRESST] is too taken aback to provide much consolation. Long since having given up on speaking truth to power, [CRESST] has the modest expectation that research should contribute to reasoned debate about the effects of state assessment policy on school practices [but] the spectacle is far from over.

"In November 1994, Lisa Graham won the election as Arizona Superintendent of Public Instruction...replacing staff with backgrounds in teaching and curriculum with people experienced in the private sector.

"With no advance warnings and no expert or public debate...Graham announced that the [Arizona] performance test was 'suspended.'...Design teams of teachers, business leaders, and parents (but no curriculum specialists) were commissioned to write standards....Hearings would then be conducted around the state...."

[CRESST thought these actions should be dropped in favor of "further research."]

"[Governor Symington signaled] to his appointees on the Board [of Education] to follow through on his conservative agenda and not give in to the professionals....[in] the confluence of Symington's bankruptcy and criminal indictments, his plan to seek reelection anyway, Keegan's own gubernatorial ambitions, and her open criticisms of his administration and calls for his resignation. Word was, he had even tried to lure her out of town to a governors' meeting on education that coincided with the board meeting."

At this point, the EducationNews reader might wonder what these biased, pompous, and catty comments, culled from four reports in the CRESST series on Arizona testing policy, have to do with the testing research we all thought we were paying for. Here's the answer:

"...mandated assessment programs are more than marks on optical scanning sheets, assignment of rubric scores to essays....One must examine instead the dynamics of wins and losses in the political arena."

Is this the kind of "research" that our U.S. state elected officials want at the Education Commission of the States? I doubt it. As I wrote to one ECS staffer several weeks ago, do the governors know what the

BEST COPY AVAILABLE

[ECS] staff is doing?

On the second of the two prevalent CRESST themes-that tests are VERY expensive; so expensive that you probably don't want to use them--two CRESST reports are most pertinent, numbers 276 and 491, though at least several others deal with the topic tangentially.

In report 276, CRESST performed a benefit-cost analysis of a new basic literacy test for teachers in Texas (the TECAT), that contained many arbitrary inclusions or exclusions of benefits or costs (see Phelps, 1996; Shepard, 1987). For example, CRESST counted the dismissal of teachers found to be illiterate as a benefit, because students would then be taught by the literate teachers who replaced them. However, in the fine print, one discovers that CRESST decided that "non academic" teachers shouldn't be counted in the benefit calculations. Which teachers were "non academic?" -- kindergarten, music, art, ESL, industrial arts, business education, physical education teachers and counselors. No matter that the citizens of Texas wanted those teachers to be literate; CRESST decided they didn't need to be.

CRESST also miscalculated the value of time by counting the benefit of the dismissed teachers for only one year, even though they were dismissed for good and the benefits would string out years into the future.

CRESST also counted costs of teachers' time spent studying for the tests, but no benefit to that studying, as if the teachers learned nothing by studying. Finally, while CRESST alleged many costs, it counted only that one benefit, from replacing illiterate teachers. There are at least several others.

After this exercise in maximizing costs and minimizing benefits was complete, CRESST declared that the Texas Teacher Test cost the citizens of Texas $53 million. Just adjusting for the mistakes in its own calculations changes the "net present value" to a positive $333 million. That's without even starting to add the benefits CRESST never mentioned.

The economists Lewis Solmon and Cheryl Fagnano estimated the values of two major benefits ignored by CRESST: the long-term labor-market benefits resulting from students learning more from more able teachers; and the attraction to the teaching profession of more able applicants as a result of higher professional standards. (Solmon and Fagnano, 1990) They estimated these benefits to be as large as a billion dollars in present value. In yet another study, the economist Ronald Ferguson found teachers' TECAT score to be the strongest predictor of Texas' minority students' success in school,

stronger than any background variable. (Ferguson, 1991)

The main theme of CRESST report 276 was that Texas had no business testing or firing teachers [tasks that should be left to "experts" and "educators"], and that the TECAT was an insult to educators and a miserable waste of time, void of any redeeming characteristics..

In another benefit-cost study, report 491, CRESST attempted to estimate both local and state-level testing costs in Kentucky and Vermont.

Among the categories CRESST actually used to classify time use by teachers were: "Preparing materials related to the assessment program for classroom use," and "working with students specifically on assessment-related tasks." The meaning of "related to," however, does not manifest strict boundaries. One activity "related to" an assessment program might be instructions in test-taking provided by a teacher just prior to a test administration. That is, indeed, an activity that would not occur if there were no test and does, then, represent a cost of the test.

Another activity "related to" the assessment, however, might be regular, ordinary instruction in a subject matter several months prior to the exam on a topic that will be covered in the exam. Both of these two activities are "related to" the assessment, but only one is a cost incurred because of the assessment. One can presume with little risk that regular, ordinary instruction existed in the period of time before the test was introduced.

Particularly in a comprehensive instructional program like Kentucky's, which attempts a "seamless" connection between curriculum, instruction, and assessment, where each element in the program mutually determines the others, teachers may not distinguish clear boundaries between each element, and may feel no caution about implicit double counting. Moreover, since all elements of the comprehensive program were implemented at the same time, teachers might readily classify curriculum or instructional activities as part of the "assessment program;" they are all component parts of only one single "program." Not surprisingly, given the ambiguity of the category, CRESST confessed to find "...a substantial range in the estimated number of hours spent by teachers in material preparation and training."

CRESST found a median number of hours "working with students on matters related to..." the testing program to be 37.4 hours per year. Hours devoted to "preparing materials related to the testing program" were only slightly less.

CRESST, moreover, also counted classroom teachers spending a median "18 hours per year administering the test" in Kentucky despite the fact that state personnel from the Kentucky Education Department actually administer the test and, in four of the five grade levels tested, the total suggested administration time is less than seven hours. This begs the question of which teacher activities CRESST classifies as "test administration."

Not surprisingly, CRESST ended up with huge estimates for the cost of Kentucky's tests. Using their base estimates, one can calculate per-student cost estimates of between $848 and $1,792. That's a lot higher than SAT or ACT prices of $20 a student (for their paper-and-pencil versions), or Advanced Placement (AP) exam prices of $65 per student. The companies producing these exams, the Educational Testing Service and American College Testing, must cover all of their costs in their prices, or they would go bankrupt. Moreover, AP exams include subjects with mostly performance-based response formats that are individually-administered and group-scored, such as their arts and music exams, just like the Kentucky exams. Nonetheless, ETS still charges only $65 an AP exam. This begs the question, is Kentucky incompetent in test administering tests, or did CRESST over count? I think the latter.

In its defense, CRESST's public relations director would probably say that individual CRESST reports are written independently by independently-minded researchers, and are independent of any CRESST group point of view. Yes, but it is CRESST's directors who decide which "researchers" get our money, and the biases of the researchers who write on policy issues have in every case been clear and obvious beforehand.

REFERENCES

Farkus, Steve, Jean Johnson, and Ann Duffett. (1997) ***Different Drummers: How Teachers of Teachers View Public Education***. New York: Public Agenda.

Ferguson, Ronald F. (1991) "Paying for Public Education: New Evidence on How and Why Money Matters," ***Harvard Journal on Legislation***, 28(2), 465-498.

Phelps, Richard P. (1996) "Test Basher Benefit-Cost Analysis," ***Network News & Views***, Educational Excellence Network, pp.1-16.

Solmon, Lewis C. and Cheryl L. Fagnano. (1990) "Speculations on the Benefits of Large-Scale Teacher Assessment Programs: How 78 Million Dollars can be Considered a Mere Pittance," ***Journal of***

Education Finance, V.16, N.1.

Post your comments on this article

EducationNews.org
The World's Leading Source of EducationNews
May 22, 2001

The Engines of Our Ingenuity
KHUF

CALL FOR PAPERS
In Defense of Testing

Test Bashing Series

Intellectual Property and Digital Information in Higher Education

LoneStar Foundation
Public Education Reform in Texas
Comprehensive Progress Report

Shop Here for Books
amazon.com

About Us

Advertising

Archives
DailyNews
Past Five Days

Book Reviews

College Aid

Book Reviews

College Aid

Test Bashing, Part 14

The "New" Rand Report: Of Course, It's Biased and Contrived, and You're Surprised?

by Richard P. Phelps

A small group of researchers at the Rand Corporation whose two most lucrative sources of contracts will likely be threatened if George W. Bush is elected president, decided recently to conduct an analysis of the Texas testing program. They released their report two days ago, but it is pure coincidence that it was released just before the presidential election. It is also only coincidence that they chose to study Texas, rather than any number of other states with testing programs very similar to the one in Texas. As James A. Thomson, CEO of the "scrupulously nonpartisan institution" says, "Texas was studied because the state exemplifies a national trend toward using statewide exams as a basis for high-stakes educational decisions." In an unusual move, Rand paid for study itself. Think of it as a public service, from a generous corporate benefactor.

Rand could have studied the testing system in virtually any other industrialized country at any time in the past few decades, since all but a tiny handful have had popular high-stakes testing programs in place for decades. Rand also could have studied any of those programs in other states or Canadian provinces that have had high-stakes testing programs for years. But, it simply just occurred to them a few months ago, sort of out of the blue, that now was a good time to do their study and Texas was a good place, honest. Ironically, pretty much every other vested-interest anti-testing group in the country seems to have received the exact same message, sort of out of the blue, just in the past several months. So ironic it's creepy.

It's also pure coincidence, surely, that this particular group of researchers at Rand spends most of its work time on contracts with two largely overlapping organizations as outspoken in their opposition to testing for over a decade as this new Rand report is now, the taxpayer-funded Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the National Research Council's Board on Testing and Assessment. Both of these organizations conduct

Ed. Links
Ed. Org
ELA
Grammar
Hispanic Links and Resources
History
Home School
Job Postings
K-12 Science
NEW Learning Disabilities and High-Stakes Standardized Tests
Math
Media Links
Mental Health
Texas Reading Institute
No Child Left Behind
National Reading Panel Report
Phonemic Awareness
Reading Kits
Reading Tests
Ideas Practicas Para Padres de Familia pdf file
Special Education Resources
TAD for English/Language Arts/Reading
LA Times Special Report Special Education a Failure on Many Fronts

research consistently supporting the status quo structure of education in the United States, from which they themselves benefit greatly, not unlike the Tobacco Institute's research program on smoking, and with similar results.

This new Rand study faced a number of challenges from the start. First, there was the bothersome detail of two other recent, and very thorough, studies by Rand personnel that extolled the virtues and successes of the high-stakes testing programs in Texas and North Carolina. These studies were conducted by independent researchers, with no ties to any education interest groups, from a different part of the Rand organization.

This new study (hereafter, "NewRand") refers to just one of them and uses the standard finesse that that report claimed "more research was needed" to discount its findings. (Reports written by honest researchers tend to make such disclaimers.) If one picks up those two studies, by David Grissmer and others, however, one for the National Education Goals Panel and the other published under Rand's banner, one will detect little equivocation. The evidence of the success of the Texas testing program is seen in the Grissmer reports to be strong and convincing.

Second, there was the bothersome detail of the rather obvious success of the Texas program, as verified, for just one example, by the coincidental and substantial improvement of Texas students on the independent National Assessment of Educational Progress (NAEP). As NAEP score increases represent the least controvertible evidence that the achievement gains in Texas are real, this NewRand report chose to attack this evidence in particular.

One really has to admire these Rand researchers; it's not easy to fabricate a rationale for declaring the obvious to be false, but these are clever people and, darn it, they found a way.

The NewRand report supports its case on four pillars:

1) While there has been an improvement in fourth-grade NAEP mathematics scores in Texas over time, there has been no improvement in eighth-grade math scores or fourth-grade reading scores.

2) What improvement there is in math and reading does not last past fourth grade. Between the fourth and eighth grades, the gain in scores over time is no greater than the average for the nation.

3) Because Texas Assessment of Academic Skills (TAAS) scores have improved by a greater proportion than Texas' NAEP scores have

Parental Opt-Out Form

Public Policy Org

NEW Publications francaises et autres sources de media

Spanish Instruction Tips

Speakers Conferences

Stossel in the Classroom

Reading

Technology

Test Bashing Series

Hot Topics & Research

The History of Education and Childhood

Fool's Gold: A Critical Look at Computers in Childhood

Block Scheduling Research

Class Size vs Performance

National Center for Alternative Certification

Sex-Ed Who's in Control

Year Round Education

Public school surveys

Texas Districts of Interest

TexasTests.org

improved, the TAAS scores must be "inflated" and unreflective of "real" gains in achievement.

4) "Evidence" from other studies of the TAAS supports the NewRand conclusions.

Here's an illustration of what the NewRand researchers were up against. Of all the states that have participated in all the state-level NAEP assessments in math and reading in the 1990s, only one other state, North Carolina, has improved its scores more than has Texas. North Carolina, of course, tests its students even more often than Texas does, and for high-stakes.

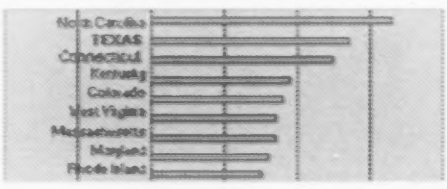
(Why haven't the anti-testing groups been visiting North Carolina this past year? Beats me. It couldn't be because North Carolina's testing program was inaugurated and has been run throughout its history by Democratic administrations, could it? No, that's too cynical an explanation. Despite the eerie similarity between the North Carolina program, inaugurated by former Governor (and now Al Gore advisor on education issues) James Hunt, and the Texas program, it must just be pure coincidence that each of the anti-testing groups, sort of out of the blue, chose to find fault with Texas instead.)

If one simply adds up the scale-score gains (or losses) over time from the different NAEP administrations for each state, one finds these results: North Carolina increased by 33 scale points overall, Texas by 27 points, and Connecticut by 25 points. These top three states all test their students a lot. In Connecticut, high stakes are not attached for the students, but the state education department uses the test information to evaluate schools and districts in a rigorous manner. Connecticut's education department is as intrusive in local affairs as many European national education departments; their quality monitoring being as thorough and intensive.

After the top three states, the cumulative scale score gains drop to 19 in Kentucky (another state with lots of testing) and on down to -10 in D.C.

Figure 1 presents the situation.

Figure 1: New Rand report says Texas no better than average -- Net cumulative scale score gains on State NAEP in the 1990s, by state



BEST COPY AVAILABLE

Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998

Statewide Mathematics Assessment in Texas

Grade 1 Reading programs as approved for 2000 local Texas adoption

Fort Bend

Notice and Declaration of Parental Rights (NDPR)

Plano

Got too much on your plate?

Texas' net cumulative score gains are more than twice the national average, yet, no matter how hard the NewRand group tried to find an improvement in Texas' NAEP scores, they just couldn't find it. In their words, they "generally found only small increases, similar to those observed nationwide, in the Texas NAEP scores." You can see for yourself in figure 1 that Texas' score increases are no different than the national average. Look. OK, look again. Don't see it, yet? Me neither. Must be our eyesight.

NewRand finds Texas' gains to be no different than the Nation's by separating the big picture above into several smaller pictures and then relying on statistical-testing artifacts within each. This is invalid, of course, because the *conclusion* they make refers to the big picture, but they don't conduct a statistical test on the big picture. They could, but they don't.

Instead, NewRand looks at a segment of gains in fourth-grade math, a segment of gains in eighth-grade math, and so on. With each segment they conduct a statistical test that relies on arguably standard, but still arbitrary, cutoff thresholds to determine "statistically significant" differences. For each separate case in isolation, there's nothing wrong with this.

The NewRand researchers probably noticed, however, that for the segments in which the Texas gains don't reach the cutoff points, they just barely don't make it. The Texas gains in the case of every segment are large by normal standards of "large," just not large enough in each and every segment to make the cutoff point for the statistical test NewRand chooses to use in each case.

If one combines the various segments (in statistical jargon, this is called "pooling"), as in figure 1, however, one can both increase the statistical power of the test (by increasing the sample size) and conduct the correct test, that for the NAEP performance of Texas as a whole, rather than for separate, discrete little bits. Combining separate

tests, or subtests, at the same level of difficulty, even on different subject matter, is done all the time when identical scales are used. Witness the many studies that use SAT combined (verbal + math) scores in their analysis.

I should add, however, that I refer to comparing scores on 4th-grade scales to scores on 4th-grade scales in later years and scores on 8th-grade scales to scores on 8th-grade scales in later years. It would not be the same to compare scores on 4th-grade scales to scores on 8th-grade scales.

If the NewRand researchers want to say it is invalid to do this, even though only identical scales are used in comparison, they would also have to admit, if they were honest, that their proclamation that Texas' performance as a whole is just average is not supported by the separate little tests they used. At best, they have a "right" to make claims only about each separate little bit, not the whole. It is tempting to assume that NewRand broke the entire sample into smaller little bits deliberately so that the statistical tests would have less power. But, *that* would be a *cynical* assumption.

Texas Improvements Don't Last Past Fourth Grade

NewRand's next analytical aside is an offshoot of the first. They, again, admit that the year-to-year gains of Texas 4th-graders on the math NAEP are substantial. Then, they look at the change in scores between 1992 and 1996 of the same cohort of students, 4th-graders in 1992 and 8th-graders in 1996 in math and 4th-graders in 1994 and 8th-graders in 1998 in reading, and find the scale-score increase to not be statistically significantly different than the U.S. average increase.

In comparing scores on 4th-grade scales to scores on 8th-grade scales they rest on much shakier ground than before. As far as I can tell, no effort was made by NAEP to make the scales comparable.

The Third International Mathematics and Science Study (TIMSS) is in the process of doing this sort of analysis, comparing a 4th-grade cohort on the its 1995 test with the same cohort (but, not the exact same students) as 8th-graders in 1999. In their case, TIMSS has made the effort to make the scales comparable by, for example, using some of the same items from the 4th-grade test in the 8th-grade test in order to jointly calibrate the two scales.

NewRand has not made any such effort. In an earlier study, in which the TIMSS compared 4th-grade to 8th-grade scores, both from the 1995 TIMSS, they wrote:

"It is important to remember that the fourth- and eight-grade scales are not directly comparable. For example, it is not the case that the eighth graders in Singapore outperformed the fourth graders by 18 points, nor is it true that fourth graders in Korea outperformed eighth graders by 4 points. [Singapore's 8th grade score average was 18 points above its 4th grade score average; Korea's 8th grade score average was 4 points above its 4th grade score average.]"

In keeping with the low- or non-comparability of the scales, the TIMSS 4th- to 8th-grade comparison was more rudimentary and conservative than the NewRanders' rather detailed analysis which pretends that the 4th and 8th grade scales are identical.

Even if one accepts the NewRand 4th to 8th grade comparison as legitimate, their analysis still retains their penchant for making proclamations about the whole after only looking separately at the parts, which is invalid.

Moreover, there remain five same-grade year-to-later-year segments in math and reading, in the 4th and 8th grades, distributed among the years 1990, 1992, 1994, 1996, and 1998. Change scores from these five segments are summarized in figure 1. NewRand now looks separately at two more, different segments representing a combination of grade-to-later-grade and year-to-later-year comparisons. In doing so, they ignore much available information but the five segments mentioned above. Moreover, they have only two segments to work with. That's not very much. If one were to add them into the mix in figure 1 the difference between the Texas gains and the nation's gains would end up exactly the same.

For an illustration of the problem, look at figure 2. NewRand looks at one segment, the second one identified in the legend, in pink, showing a change between year 2 (1994) and year 4 (1998).

While in this focus, they ignore information to be gained from the other two segments, year 1 to year 2 and year 4 of the year-to-year comparisons. They see an improvement of just 1 point and miss an improvement of 4 points. They just have too little information to be making summary judgements.

BEST COPY AVAILABLE

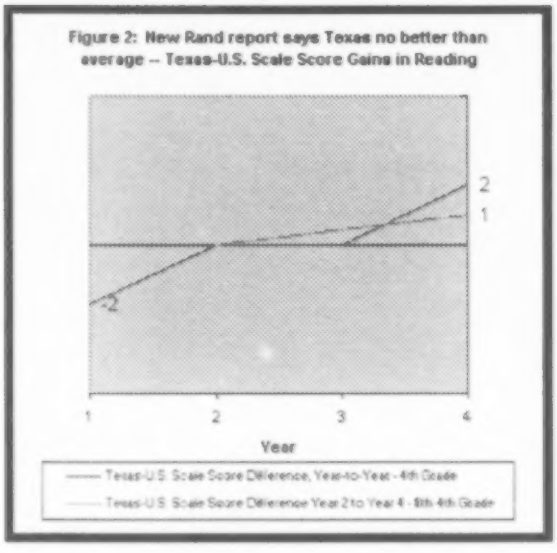


Figure 2: New Rand report says Texas no better than average -- Texas-U.S. Scale Score Gains in Reading
2
1
-2
1
2
3
4
Year
Texas-U.S. Scale Score Difference, Year-to-Year - 4th Grade
Texas-U.S. Scale Score Difference Year 2 to Year 4 - 8th 4th Grade

The situation is even more serious in math, where the NewRanders would throw out some of the country's largest scale-score gains of the 1990s -- 11 points in 4th grade math between 1992 and 1996 and 12 and 6 points in 8th grade math between, respectively, 1990 and 1992 and 1992 and 1996.

No, no, says NewRand. A conclusion about the whole is made based on separate tests of the parts, each of which must, separately, be statistically significant, or we must conclude there to be no difference whatsoever between Texas and the national average for the whole.

Using their methodology, probably no state in the country can be found to be any different than the national average. Remember, Texas is the second most improved state in the U.S. adding NAEP year-to-year measures, with more than twice the improvement than the average, and NewRand says it is no different than average overall. Unless NewRand wants to assert that it is not valid to compare scores on the same test year to year, which they are not about to do without being ridiculed, their conclusion of no difference between Texas and the U.S. must be wrong.

They're missing the forest for the trees.

TAAS Gains are Disproportional to NAEP Gains, and So Cannot be "Real"

While the two NewRand analyses discussed above could *arguably* fit into the murky category of misunderstanding or disagreements over methods, or some such, the next argument is nothing more than gross misrepresentation.

BEST COPY AVAILABLE

Scores on two tests cannot be perfectly correlated without them being the same test. The TAAS and the NAEP are not the same test, nor are they supposed to be, so their scores will never be perfectly correlated.

The fact that the increases in the TAAS are greater than Texas' students' score gains on the NAEP are to be expected. Any other result would reveal a serious problem. The TAAS contains subject matter that matches the curriculum standards of the state of Texas. The NAEP does not. Therefore, it is to be expected that Texas student scores on the TAAS will increase more than Texas student scores on the NAEP.

Testing opponents' cry of "teaching to the test" is obfuscation. Teaching to the test is only a problem when students are tested on material they have not been taught. When students are tested on material they have been taught, any teacher not teaching to the test is behaving irresponsibly. It may be for this reason that, despite testing opponents' (largely successful) efforts to convince journalists that teaching to the test is a horrible practice, parents continue to tell pollsters that, of course, they want their students' teachers to teach to the test.

There is nothing sacred about the content of the NAEP. It no more represents "real" learning than the TAAS does. The NAEP tests a reasonable, but ultimately arbitrary, sample of academic subject matter. It is based on no legal curriculum and NAEP content does not match the curricular standards in any U.S. state.

This is not to say that the NAEP is not a good test; it is. But, there is no sensible reason why performance on the TAAS should exactly match performance on the NAEP. The fact that Texas' students' score gains on the NAEP have been second best in the country, provides evidence that the improved achievement of Texas' students is generalizable beyond the confines of Texas' curriculum. It can in no reasonable way be argued to be a failing.

If the NewRanders want to continue this line of reasoning, claiming that the TAAS must be deficient because its gains don't precisely mirror the NAEP's, they would also have to admit, if they were to be honest, that the situation must be *worse* in every other state in the country, other than in North Carolina. Ultimately, that is where their argument leads us if we assume that the NAEP should represent a mirror-image barometer of state test outcomes. The NewRandies would have to admit that, after North Carolina, Texas is the *least worst* state in the country on academic achievement gains. Relatively, even according to NewRand logic, Texas still ends up second best in the nation.

"Evidence" from other studies of the TAAS supports the

NewRand conclusions

How low will the vested interests in education go in order to protect their little fiefdoms? Apparently, very low, indeed.

With this fourth of the NewRand arguments, our special group of Rand scientists dives headfirst into the muck with the sleeziest of the sleeze, and their CEO may follow with the formerly good name of the Rand Corporation.

As verification of their work, the NewRanders mention the similar conclusions found by other testing-opponent researchers. Either they haven't read the "research" they cite, or they couldn't care less whether any of it is accurate. I can believe either. The NewRanders cite as evidence "facts" about the Texas testing program that are easily shown to be results of extraordinary research errors (accidental or otherwise) and, in some cases, complete fabrications.

For example, the NewRanders write:

"It is worth noting that even the relatively small NAEP gains we observed might be somewhat inflated by changes in who takes the test. As mentioned earlier, Haney (2000) provides evidence that exclusion of students with disabilities increased in Texas while decreasing in the nation, and Texas also showed an increase over time in the percentage of students dropping out of school and being held back. All of these factors would have the effect of producing a gain in average test scores that overestimates actual changes in student performance."

Every statement in this quotation can easily be shown to be false. Haney found Texas to be excluding more LEP and disabled students from the NAEP exams only because he (accidentally or deliberately) misread a table in a NAEP report. The reality is just the opposite. Texas is among the nations' leaders (1st at one grade level and 4th at the other) in LEP NAEP test inclusion among states with large populations of limited-English-proficient (LEP) students. In truth, Texas actually made it harder on itself by going out of its way to include more LEP students than did most states. Had Texas included the same proportion of LEP students in its NAEP testing as did the average state, Texas' NAEP scores could have risen even higher than they did.

Haney thinks otherwise because he compares Texas's percentages from the main NAEP sample to averages for the nation from the completely different national tracking sample. Haney's erroneous analysis would show every state in an unfavorable light, not just

BEST COPY AVAILABLE

Texas, if he were to apply it to the other states.

For more, see:
http://www.educationnews.org/test_bashing_part_9.htm

Haney finds Texas to be high in dropouts only by changing the definition of what a dropout is without clearly telling the reader he's made the change. By his definition, the dropouts rates in all states would be higher than they really are. But, of course, he doesn't do that. He surreptitiously changes a definition, applies it to Texas, and uses it in comparisons as if it were the official number. This is one of the several ways he artificially gets Texas to look bad.

In fact, Texas has a relatively low rate of dropout and a relatively low rate of grade retention

If one compares Texas' dropout rate (the real one, not Haney's contrivance) to those of its demographic peers, the other Southern or Mexican Border states, Texas comes out above average. Probably the most pertinent demographic factors affecting dropout rates are the percentages poor and minority of a student population in a state. It is, indeed, unfortunate, that Black and, particularly, Hispanic students tend to have higher dropout rates, but they do. And, Texas has a much larger Black population than do most states. Texas also has the second largest population of Hispanic students in the country, second only to California. Moreover, Texas's minority population tends to be poor. Given these demographic challenges, it is remarkable that Texas has held down its dropout rate to less than that of other Southern and Southwestern states. Texas educators should be commended for such success, not criticized.

The converse of a dropout rate is a "persistence" rate, the proportion of students who remain in school year over year. Haney says that Texas' rate of persistence is remarkably lower than that of other states. Haney, however, takes liberties with the definition of a persistence rate, too. He compares grade 9 enrollments to grade 12 graduation numbers. He didn't just pick grade 9 out of a hat; the reason he likes using grade 9 is explained below. There's also a reason he looks only at grade 12 graduation numbers, rather than graduation numbers for an entire cohort; that way he can make it look like more Texas students are dropping out.

I compare both grade 8 and grade 9 enrollments to grade 12 graduation numbers. Texas has a higher-than-the-national-average persistence rate and a rate higher than those for other Southern and Southwestern states, save one, for the grade 8 comparison.

For the grade 9 comparison, Texas is hobbled by a large bulge of

students retained at grade 9 and, not surprising, its persistence rate slides under the national average this time. However, compared to its peer states, Texas is still above average, higher than six out of the nine with full data available.

For more, see:
http://www.educationnews.org/test_bashing_part_11.htm

Haney writes about grade 9 as is it was a magic grade, much, much more important than all the others. It just may be that he harps on grade 9 because it is there that Texas retains a high proportion of students (17%) and he wants to make a point that Texas is onerous in flunking students (and that retaining a student scars him for life and makes him want to drop out of school). Haney has no tolerance for retention by any rationale. Whether a student studies or not, whether a student learns anything or not, whether a student shows up at school or not, all students should be given high school diplomas "on time" regardless, according to Haney.

Does Texas retain students in grade at a rate much higher than other states, as Haney claims? Absolutely not. Indeed, grade 9 is the only grade where Texas's retention rate is above the average (except for grade 1's 0.1 difference) for nine states with complete data on retention. No wonder Haney picked grade 9. Any other grade, 2 through 8, and Texas looks like a wimpy social promotion state, the kind that Haney reveres.

Texas comes in below the average state sum for retention rates in grades 1 through 9, ranking 7th in cumulative retention rate out of 9 states. A look at the upper grades shows much the same picture. Texas ranks below average on retention for grades 10 and 12, and slightly above average for grade 11. Its cumulative retention rate for all grades 1 through 12 is well below average, ranking 6th out of only 9 states in retention.

Texas ranks very low on retention.

For more, see:
http://www.educationnews.org/test_bashing_part_11.htm

Genuine, reliable, and accurate data on LEP exclusions from NAEP tests, dropouts, grade retention, and persistence rates show Texas to be better than average on all measures. Because this fact does not square with Walt Haney's preconceived bias, he makes up his own measures and, in some cases, he just pulls numbers out of thin air. His analysis is a rickety construction that can't hold up against a couple of hours of casual scrutiny. His alleged evidence is worthless. These

people will say anything.

For more, see *Test Bashing, Part 8 through 11*, at

http://www.educationnews.org/test_bashing_series_by_richard_p.htm

There has, perhaps, never been a better argument than these tawdry exercises in advocacy research for dismembering the education research establishment, that is, if the American people are to have any hope of honest discussion of policy issues based on education research. Time and time again, "research" conducted by interests vested in the current system of education has shown itself to be untrustworthy. Most galling, as with the cushy work contracts the NewRandies are trying to preserve for themselves, we taxpayers pay for most of this "research" that is designed to fleece us.

Thinking from Inside a Tank

Anymore, most "Think Tanks" are little more than Beltway Bandits with more pretensions. The danger of losing two major lines of business, especially from two honey pots paying the extra-high salaries these NewRand "senior scientists" have gotten used to...for work performed with virtually no oversight...has got to rivet one's attention. Rich tastes, once developed, are difficult to give up.

Anyone could have sympathy for their plight, nonetheless. The fear of losing one's major sources of income with no comparable replacement in sight, while one's children are in college and the mortgage still isn't paid off...can have a deleterious effect on one's dedication to objectivity.

It's not likely the NewRandies will be able compensate for CRESST and NRC Board work if those gravy trains stop running. It's well-paid and easy work; they can write pretty much whatever they please without any significant review. But, even if George W. Bush had no intention of stopping these gravy trains before, he certainly might try to now if he gets elected, and he certainly should. The NewRandies might then be forced to find real work like the rest of us. Tis a pity.

The Free Press to the Rescue

But, really, do the NewRandies have anything to fear? The Press, after all, has attached itself to their story with a strong suction, in keeping with the rest of its coverage of testing issues this campaign season. Outside the small, but well-funded and vocal, universe of those who directly profit from status quo recalcitrance in education, U.S. journalists represent the only other group in the country feeling antipathy toward fair consideration of the testing issue.

I have downloaded several hundred newspaper and magazine articles and TV show transcripts on testing from the Web to see who they relied on for "expert" commentary. I have not seen a single instance, out of dozens from the past several months, where an "expert" with a favorable point of view toward high-stakes testing was interviewed. Not one. Meanwhile, in each of those dozens of articles or TV shows which featured expert interviews, the "expert" or 'experts" interviewed were well-known opponents of testing, usually one of the country's most extreme.

As for Texas in particular, there was many months ago a lawsuit against Texas' test, the TAAS, alleging it to be discriminatory. Expert witnesses spoke for both sides and the suit eventually failed. I have since seen many articles and transcripts featuring interviews with prosecution expert witnesses, those opposed to the Texas test. I have yet to see one that interviewed any of the defense expert witnesses.

Though I have not yet had the time to actually count up column inches, I will not be surprised if, when I do, I will find that only one article or TV show from the past few months that interviewed pro- and anti-testing advocates of any kind gave equal time to both sides. (Bill Kurtis of A&E's "Investigative Reports" may be the only journalist in the country who has bothered with this antiquated journalistic nicety.)

If there has been any issue since World War 2 on which the press coverage has been more one-sided, I don't know about it. In the face of decades worth of overwhelming public, parent, student, and teacher demand for higher standards, more accountability, and more testing, our country's journalists seem to have chosen, as a pack, to side with a small group of fat cat reactionaries who profit from the status quo, and to hell with the greater good.

Richard P. Phelps is co-author, with Barbara Lerner and Gregory Cizek, of *The War on Testing*, forthcoming from Lawrence Earlbaum. He has also written "Why Testing Experts Hate Testing," which will soon be available in both English and Spanish:

http://www.edexcellence.net/library/phelps.htm

1. For more, see *Test Bashing, Part 13*, for CRESST, at

http://www.educationnews.org/test_bashing_series_by_richard_p.htm

and http://www.siop.org/tip/backissues/Tipapr99/4Phelps.htm

for the NRC Board. Another review of another NRC report will appear in the next issue of *Educational and Psychological Measurement* (after the

election).

2. Mullis, I.V.S., et.al. *Mathematics Achievement in the Primary School Years*, TIMSS, p.40. In a discussion on the same topic, the Organization for Economic Co-operation and Development (OECD) wrote this caution: "Since there were relatively few items in common, the size of the link is approximate and the achievement increases between 4th and 8th grades must therefore be interpreted with caution." OECD, *Education at a Glance: OECD Indicators 1998*, pp.314-315. I was involved in performing this analysis for the OECD and, I can assure the reader, we did not just blindly accept the two scales as equivalent, but performed a considerable amount of analysis to see how they related to each other.

3. Haney defines a dropout to be someone who does not graduate "on time," exactly 12 years after they enter primary school, even if they do graduate, even if they graduate just a month later. No responsible statistical authority anywhere defines the term this way. Using the Haney definition, dropout rates become far higher than official dropout rates. That is because some students do not graduate "on time." Many students everywhere, not just in Texas, graduate "late" for a variety of reasons, such as being retained in grade somewhere between kindergarten and graduation, failing a course in 12th grade (and making it up in summer school), missing a semester due to illness, making up requirements after having moved from another state with different academic requirements, and so on.

Post your comments on this article